

AD-A054 650

ROME AIR DEVELOPMENT CENTER GRIFFISS AFB N Y

F/G 14/2

PROCEEDINGS OF THE RADC SPECTRUM ESTIMATION WORKSHOP HELD ON 24--ETC(U)

MAY 78

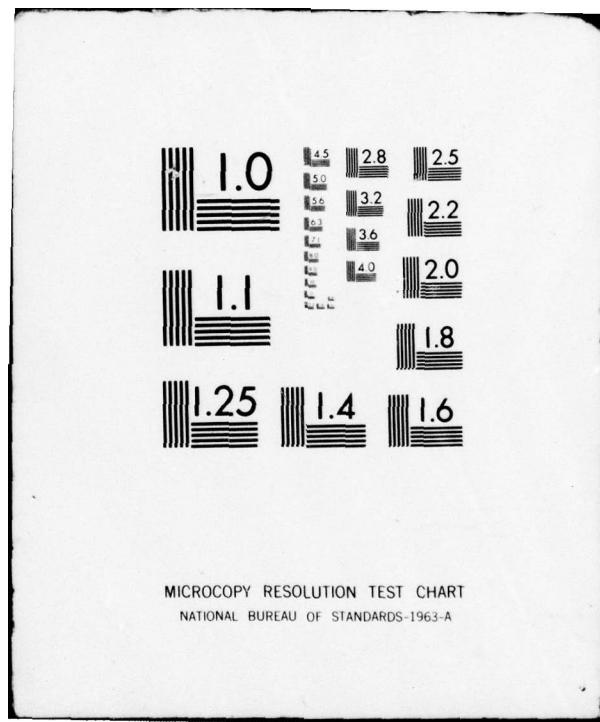
UNCLASSIFIED

NL

1 OF 4

AD  
A054650





AD A 054650

FOR FURTHER TRAN ~~DU~~

PROCEEDINGS OF THE

RADC

*Spectrum Estimation*

WORKSHOP Held

On 24, 25 & 26 MAY, 1978.

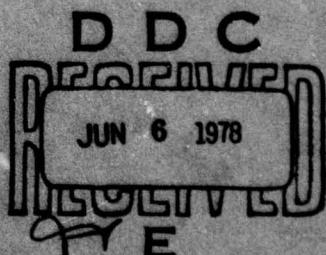
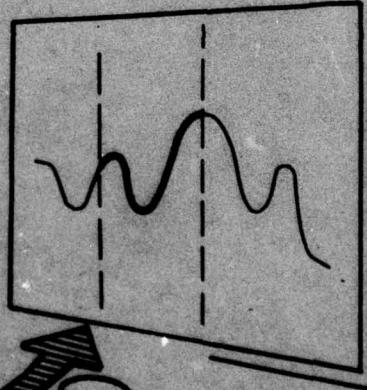
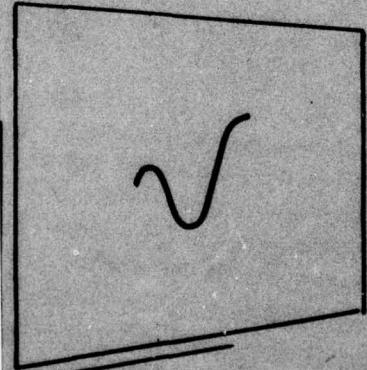
act 2

11 May 78

12 294 P.

Approved for public release; distribution unlimited.

DDC FILE COPY



309 050

LB

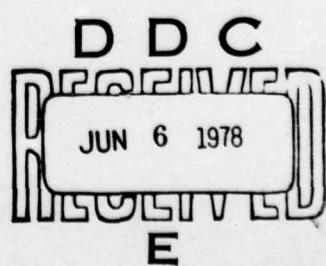
This report has been reviewed by the Office of Information, RADC, and is  
releasable to the National Technical Information Service (NTIS). At NTIS  
it will be releasable to the general public, including foreign nations.

RADC's SPECTRUM ESTIMATION WORKSHOP

AGENDA

MAY 24, 25, AND 26 1978

Wednesday, May 24		Page
0800	Registration	
0900-0915	Welcome - Colonel Owen R. Lawter Chief, Surveillance Division	
0915-0945	RADC Overview - Dr. John Burgess Chief Scientist	
0945-1015	Approaches to Spectral Analysis - New and Used, Dr. Lester Gerhardt (Co-chairman), RPI	1
1015	Coffee Break	
Session I	Lt Colonel William Cuneo, DARPA	
1045	Frequency Resolution of High-Resolution Spectrum Analysis Techniques, Larry Marple, Signal Science, Inc.	19
1105	Autoregressive Model Spectral Estimation, Some Simulation Study Statistical Performance Results, Will Gersch and R.Z. Liu, University of Hawaii	37
1125	High Resolution Spectral Estimation via Rational Models, Mostafa Kaveh, University of Minnesota	51
1200	Adjourn for lunch	
Session II	Dr. Henry Radoski, AFOSR	
1330	Maximum Likelihood Spectral Estimation Using State-Variable Techniques, Robert McAulay, MIT/Lincoln Laboratory	63
1350	An Optimum Filter for Spectral Estimation Based on a Penalty Function Approach, R.J.P. deFigueirido, Rice University	69



1410	A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis of Complex Signals, Paul F. Fougerre, AF Geophysics Laboratory	77
1430	Coffee Break	
1500	Adaptive Extrapolation and Hidden Periodicities, A. Papoulis and C. Chamzas, Polytechnic Institute of New York	85
1520	Recursive Spectral Estimation of $\alpha$ - Stationary Processes, Dr. M. Morf and D. T. Lee, Stanford University	97
1540	Improved Spectral Estimation From Incomplete Sampled - Data Observations, James A. Cadzow, Virginia Polytechnic Institute & State University	109
1615	Adjourn for the day	

Thursday, May 25

Session III	Dr. Donald Burlage (DRDMITER) U.S.A. R&D Missile Command	
0830	A Review of Prony's Method Techniques for Parameter Estimation, Michael L. Van Blaricum, Mission Research Corporation	125
0850	System Identification by Use of Pencil-of-Functions, V.K. Jain, University of South Florida; D.D. Weiner and J. Nebat, Syracuse University; T.K. Sarkar, Rochester Institute of Technology	141
0910	Two Dimensional Spectral Estimation, Anil K. Jain, Surendra Ranganath, State University of New York at Buffalo	151
0930	Lattice Methods in Spectral Estimation, John Makhoul, Bolt Beranek & Newman, Inc.	159
1000	Coffee Break	
1030	Non-Parametric Spectrum Estimates Motivated by the Wishart Distribution, David J. Thomson, Bell Telephone Laboratories	175

1050	Air Vehicle Detection Using Advanced Spectral Estimation Techniques, Philip G. Tomlinson, Guy A. Ackerson, Decision-Science Applications, Inc.	191
1110	Radar Imaging of Discrete Targets with Maximum Entropy Techniques, Stephen B. Bowling, MIT/Lincoln Laboratories	207
1200	Lunch	
Session IV	Dr. Gerard Trunk, Naval Research Laboratory	
1330	Application of Maximum Entropy Frequency Analysis to Synthetic Aperture Radar, Philip L. Jackson, Lawrence S. Joyce and Gerald B. Feldkamp, Environmental Research Institute of Michigan	217
1350	Antenna Patterns Computed with Maximum Entropy and the Burg Technique, William R. King, Consultant, Alexandria VA	227
1410	Maximum Entropy Cepstral Analysis, Tom Landers, MIT/Lincoln Laboratories	245
1430	Coffee Break	
1500	A New Adaptive Filter for Radar Clutter Rejection, D.E. Bowyer, P.K. Rajasekaran, W.W. Gebhart, Teledyne Brown Engineering	259
1520	Doppler Spectrum Estimation for Continuously Distributed Radar Targets, George R. Cooper, Clare D. McGillem, Purdue University	273
1540	Instantaneous Frequency Estimation from Sampled Data, William R. Carmichael, Richard G. Wiley, Syracuse Research Corporation	287
1615	Adjourn for the day	

**Friday, May 26**

0830	A Comparison of Solutions to the Workshop Problems, Dr. Lester Gerhardt (Co-Chairman), RPI	
1030	Coffee Break	
1100	Workshop Panel Activity	
1230	Adjourn the Workshop	

White Section	<input checked="" type="checkbox"/>
Buff Section	<input type="checkbox"/>
BY.....	
DISTRIBUTION/AVAILABILITY CODES	
DIST.	AVAIL. and/or SPECIAL
A	

## PREFACE

This workshop provided a means for key researchers in the field to describe their work and also provided a means for comparing the work of various researchers using a common data base for representative problems of importance to the Air Force. This report is a collection of papers that were submitted for presentation at RADC's Spectrum Estimation Workshop held 24-26 May 1978 at Griffiss Air Force Base, N.Y. 13441. The papers were published as received by RADC and have not been edited. Further, publication of these papers does not represent approval or endorsement by the Rome Air Development Center or the U.S. Air Force.

The researchers were also presented with a set of sample problems called the Spectral Estimation Experiment. The object of this experiment was to establish a basis for comparison of the wide variety of techniques available as a function of selected applications on both real and artificial data sets representing specialized problem classes which are of interest to the government. The common data base offers several additional advantages.

Three different problems have been formulated by the workshop committee. They fall generally into the areas of radar, pattern recognition and system identification.

The detailed description of the problem and the solutions as determined by the many different algorithms employed will be published separately.

### SPECTRUM ESTIMATION WORKSHOP COMMITTEE

1. Edward Christopher (RADC/OCTS)
2. Lester Gerhardt (RPI/Co-chairman)
3. James Roach (RADC/DCIT)
4. Clarence Silfer (RADC/OCTS/Co-chairman)
5. Paul Van Etten (RADC/OCTS)
6. Haywood Webb (RADC/ISCP)

## APPROACHES TO SPECTRAL ANALYSIS - NEW AND USED\*

LESTER A GERHARDT

Professor and Chairman  
Electrical and Systems Engineering  
Rensselaer Polytechnic Institute  
Troy, New York 12181

### Abstract

This paper served as the introductory address for the Spectral Estimation Workshop sponsored by Rome Air Development Center in May 1978. Following a short history of the development of techniques for spectral estimation, the major approaches are summarized. These include the more traditional transform approach, the rediscovered autoregressive estimator and related maximum entropy spectral analysis (MESA) methods, and Prony's method. Also covered are new approaches including an algorithm for band limited extrapolation, among other methods. The major approaches are then simply compared with respect to data required, resolution, application, and sensitivity to noise. Finally, the papers presented at the Workshop are briefly reviewed and grouped with respect to the classes of techniques discussed.

### Introduction and Background

In the way of historical background, the "raison d'être" for this Workshop probably is best explained by the recent rebirth of the autoregressive spectral estimation techniques. The potential of such methods to more accurately estimate parameters of certain types of spectra, particularly peaked spectra characteristic of radar applications was initially viewed by many as offering "super-resolution" with capability beyond the diffraction limit (in any case better than the classical techniques in resolution for certain signal to noise ratios). As a result of investigations and applications of these methods, a broader and more generalized view towards spectral estimation evolved. This led eventually to the organization of the Workshop, for the consideration and comparison of several classes of approaches to spectral estimation by experts in the field advocating the different methods. In the process, a more realistic view of the advantages and liabilities of each of the approaches was formulated coupled with a better understanding of the interrelationship that exist among the techniques and an appreciation of the historical significance of their development and use.

\*The contributions of E. Pflug, ESE student at R.P.I., to the literature survey and technical discussions are gratefully acknowledged.

In the way of mathematical background, spectral analysis is well accepted as a tool to aid the understanding of the signals encountered in the physical world. The mathematical foundations are due to the French mathematician Baron Jean-Baptiste Fourier, who established the explicit relation between the time domain function (signal) and its unique frequency domain function (the spectrum), hence the name Fourier Spectrum. Despite the establishment of this relationship, it was more than 100 years before major applications involved the concept (1). The Fourier transform pair relating a continuous aperiodic signal with its transform is simply

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega$$

It is usually the magnitude of the transform that is termed the spectrum of the signal. More often, the signal itself is random or stochastic but stationary in nature, and can be best described by its correlation function  $R(\tau)$ . The Fourier transform of the correlation function is the power spectral density which together form a Fourier transform pair.

$$\phi(\omega) = \int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau} d\tau$$

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\omega) e^{j\omega\tau} d\omega$$

The above basic concepts apply to continuous signals and imply infinite limits; that is an infinite amount of data is needed to obtain the spectrum.

In more practical situations, the signals of interest are sampled in time thus forming a discrete time signal. Such signals are transformed to a discrete line spectrum using a discrete fourier transform (DFT). The DFT is expressed as

$$F(k\omega) = \sum_{n=0}^{N-1} f(nT) e^{-j \frac{2\pi nk}{N}}$$

where  $k = 0, 1, \dots, (N-1)$  and  $f(nT)$  are uniformly spaced samples of the waveform at intervals  $T$ .

As above, only a finite number of the signal samples are available. Thus the problem is to produce a spectrum of a finite history of the signal which best estimates the actual spectrum. It is to this end that the approaches to follow are addressed.

In many cases, continuous or discrete, the signal characteristics (spectral) change with time, that is the signal is nonstationary. Consequently, one must deal with time varying or short term spectra, a function  $F(\omega, t)$ . This may be displayed in a time varying frequency format in two dimensions. In any of the above cases there is a clear need to update the spectral estimate, either because it uses a finite number of data points, or because the characteristics are time varying or both. This need for continued recomputation of the transform led to the discovery of so called fast transforms such as the FFT (Fast Fourier Transform) which greatly reduce the computations needed and permit higher speed processing. An FFT typically reduces computations from  $N^2$  to  $N \log_2 N$ , in the order of 2% of that of a DFT for  $N = 500$ .

Some of the most influential work in developing an estimate of the power spectrum using a finite number of samples or time history of the signal is that of Blackman and Tukey (2, 3). This approach will be discussed shortly in more detail. The overall approach, which modified the original data by weighting it by an appropriate window function or weighting function, yields a result which is the convolution of the transform of the window function with the spectrum of the portion of the signal selected. Because it tends to taper the original signal values, it is sometimes called the transform and taper (TT) approach (2). This approach remained one of the most popular for about a period of ten years (in the open literature).

In the late sixties, a variety of non-transform and taper approaches came to light. These included the autoregressive method (designated AR) by Parzens (4), and the maximum entropy method (MEM) of spectral analysis (MESA) developed by Berg (5) each from slightly different perspectives and fields. In several papers these techniques are referred to in a common context attesting to their similarity (6, 7). About the same time Capon designated the maximum likelihood technique (8). In a paper in this Workshop (by McAulay) it is shown that the maximum likelihood spectral estimator corresponds to solving the same normal equations as for the MEM for an all pole model with large signal to noise ratio, whereas for low signal to noise the MEM is no longer optimum. Regardless of their similarity, one thing however is clear. These techniques, offered a different approach to spectral estimation than the more classical window approach, and these were destined to be compared.

Representative of such investigations of these new techniques, H. Akaike developed an error criterion for AR techniques in 1969 (9) called the final prediction error (FPE). He soon related this to spectrum estimation (10). In 1971, R. T. Lacoss formulated a paper comparing the MEM and MIM methods (11). In 1972, T. J. Ulrych's applied the MEM method to truncated sinusoids (12) and K. N. Berk's studied the consistency and asymptotic characteristics of the AR method (18). In 1973, Edward and Fitelson (13), Gersch and Sharpe (14) and Ulrych, Smylie, Jensen, and Clarke (15) further extended these new techniques. The list goes on, with Radoski, Foufoula, and Zawalick (16) and Jones (17) investigating the MEM and AR methods respectively. In 1976, Kaveh and Cooper showed the equivalency of the AR and MEM methods and the comparison to the TT methods and investigated their properties, with some numerical examples (2). Many others continued with

applications and further investigations. Some are given in references (19-26).

An interesting aspect to note is the diversity of fields represented by these various investigators, many of whom are contributors to this Workshop. They range from the estimation theory of the Communications and Control areas of Electrical Engineering to the Geophysical Scientists, from academicians to engineers, etc. There is no question that this subject offers the opportunity to bring together the talents of many to try to merge the knowledge of systems identification, adaptive systems, signal processing using linear predictive coding, recursive estimation, Kalman filtering, classical transform analysis, statistical analysis, computer technology, to name a few to help compare these new and old methods and evaluate their place in spectral estimation. It is to this end that this Workshop is directed.

The paper will now briefly review aspects of the major methods to be later considered by the investigators, taking care not to steal any of their thunder. Some of the strengths and weaknesses will be emphasized. Last will be a review of the papers to be presented in the light of the areas described.

#### Traditional Spectral Analysis

The most straightforward approach to spectral analysis is to supply the time domain signal to a bank of narrow band filters. The bank of filters approximates the Fourier transform in the limit. The technique in concept offers almost ideal spectral analysis limited only by practicality. Similar approaches also use adaptive filters which track major concentrations of energy in the spectrum when it is time varying such as in speech, sonar, medical signals, etc. (27). Variations of this approach use swept filter analyzers which simplify the circuitry requirements but at a sacrifice in performance. They require longer acquisition times for high resolution, better linearity with a wide dynamic range, and greater stability. A further extension of such ideas are time compression analyzers. Nonetheless the filter bank approach remains one of the most popular particularly when implemented digitally.

The resolution of such a system is limited to the number of filters (and therefore the bandwidth of each filter) placed across the frequency range of interest. In practice, with these realized digitally, it is usually the computation time which limits the effective resolution. From another prospective, the time that the filter must be exposed to the signal is inversely proportional to the filter bandwidth (the Heisenberg Uncertainty Principle). As a result, the finer the frequency resolution desired, the more time (or larger the number of samples) needed. Clearly, we are dealing with a finite time (or number of samples) which must be increased to obtain finer frequency resolution. This is a fundamental difficulty with this classical approach even if implemented digitally.

The second inherent difficulty (or degree of flexibility depending on your outlook) is the effect of the window function on the spectrum obtained.

The use of a finite amount of data may be easily interpreted as a multiplication of the original signal by a window function (viewed as a pulse in the simplest case). The result is a spectrum which is the convolution of the original spectra (which was desired) by the transform of the window function. The result may be truly very different than desired depending on the selection of the window function.

In a more exact way, a stochastic signal must be treated in terms of its correlation function as mentioned before. Let  $x(t)$  be the given observed signal of finite length  $T_m$  (3). Then

$$c(\tau) = \frac{1}{T_n - |\tau|} \int_{-(T_n - |\tau|)/2}^{(T_n - |\tau|)/2} x(t - \tau/2) \cdot x(t + \tau/2) dt$$

$$|\tau| < T_n$$

where  $c(\tau)$  is the approximate autocorrelation function at lag  $\tau$ .

Let  $d(\tau)$  be the window function; an even function of  $\tau$ . Then the modified approximate autocorrelation function is given by:

$$c_m(\tau) = d(\tau) \cdot c(\tau)$$

and the power spectrum estimate is given by

$$C_m(f) = D(f) * C(f)$$

where  $*$  represents the mathematical operation of convolution. In actuality,  $C_m(f)$  is found by Fourier transforming  $c_m(\tau)$ , (windowing of the original signal is also possible). This is commonly known as the correlation approach, and is heavily related to the periodogram approach.

This is a well researched field in digital signal processing. A good summary of window functions and their effects appears in the IEEE Proceedings, Jan. 1978 in a paper by Harris (22). The emphasis he puts on the proper selection of the index of performance is important to note. These windows must be designed to compromise between resolution and stability. This essentially breaks down to making the transform of the window  $D(f)$  block-like yet with low sidelobes. One naturally contradicts the other. It is a restatement of the uncertainty principle.

Overall, then the traditional approach is still limited by the tradeoff between time measured data and frequency resolution obtainable, and the proper selection of the window function, not unrelated questions. The advantages and steps forward here in recent years have been mostly in improvements in speed of computation using FFT's and array processing, making this technique continually appealing. It still remains as one of the most popular approaches to stationary and even nonstationary spectral analysis to date despite the drawbacks cited.

### Autoregressive and Maximum Entropy

The traditional approaches (autocorrelation, periodogram, etc.) are good when sufficient data is available (long compared to the reciprocal of the lowest frequency of interest). In many problems such as radar and seismic signal processing, this is often not the case. The more recently developed methods of AR and MEM offer the possibility of better performance on shorter records.

These new techniques are not only strongly related to each other, but to other fields as well. These include work done in the past on system identification where the objective is to model an unknown system using a linear adaptive model with adjustable poles and zeroes (28), linear predictive coding (29), among others.

The AR method uses a finite autoregression fit to the time series data and calculates the spectrum from the autoregression coefficients (as well as the error variance). A key problem is to determine the order of the autoregression to be used. The difficulty is not unlike that confronting the user in a pattern recognition problem in determining the optimum number of features to use given a finite data set (such as in the work of Foley, Webb, Gerhardt, etc.). Akaike's Information Criterion is the most popular method used to determine the order of the regression, because of its effective use of the final prediction error FPE. In essence, the AR parameters are fit by a least squares to the covariance sequence of the observed data (the sets of equations to be solved for the coefficients are exactly those used for linear predictive encoding (LPE)). When the order of the estimate is the same as the order of the model that generated the data, the parameters are known to be maximum likelihood estimates. The question as to the order of the model has been the brunt of much of the recent work by Akaike and others.

Maximum Entropy is a method credited primarily to Berg, which provides an estimate of the power spectral density which maximizes the entropy of a stationary random process from the first N lags of the autocorrelation function. It has been shown (30) that this method is in fact equivalent to fitting an AR data model to the available time series data and closely related to work appearing in the statistical literature previously. With this similarity, it is only necessary to describe one of these techniques in detail. The following parallels the explanation of AR given by Griffiths (24) which appears in numerous other references as well.

The AR method of spectral analysis may be described in terms of the whitening model shown in Fig. 1. The input data sequence  $x(k)$  is filtered by a whitening operator having an impulse response with z transform

$$H(z) = 1 - a_1 z^{-1} - a_2 z^{-2} \dots a_L z^{-L}$$

where  $z^{-1}$  has been used to denote a unit delay. As is well known

$$S_e(\omega) = S_x(\omega) |H(\omega)|^2.$$

In AR analysis, the L coefficients  $a_1, a_2, \dots, a_L$  are chosen so as to minimize the power at the filter output. Zero output power is not possible due to the fact that the leading coefficient is unity. Markel and Grey (29) have shown that minimizing the output power is equivalent to providing the flattest possible output spectral density  $S_e(\omega)$ . The appropriate equations for these coefficients are called the normal equations or correlation equations, and are given in matrix form by

$$\begin{bmatrix} r_x(0) & r_x(1) & \dots & r_x(L-1) \\ r_x(1) & r_x(0) & \dots & r_x(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(L-1) & \dots & r_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(L) \end{bmatrix}$$

where  $r_x(\ell)$  is the autocorrelation of the input data sequence at lag  $\ell$ . If the resulting coefficients provide a truly flat output spectral density, i.e.,

$$S_e(\omega) = \gamma, \quad \text{for all } \omega$$

then, combining the input spectral density  $S_x(\omega)$  may be expressed as

$$S_x(\omega) = \frac{\gamma}{|H(\omega)|^2}$$

Since

$$H(\omega) = 1 - \sum_{\ell=1}^L a_\ell e^{-j\omega\ell}$$

and, as shown by Makhoul

$$\gamma = r_x(0) - \sum_{\ell=1}^L a_\ell r_x(\ell)$$

the input power spectrum can be determined directly from the  $a_\ell$ , which are called the AR coefficients.

Of course, one cannot always expect a flat output spectrum for finite values of L, which is the order of the AR model referred to before. This occurs only if the second-order statistics of  $x(k)$  can be reproduced exactly by a sequence which is generated by filtering white noise with a filter containing no zeros and L or fewer poles (30). For this reason, AR spectral analysis is often called the all-pole-model method of spectral analysis.

The theory outlined above assumes that the data autocorrelation function  $r_x(\ell)$  is known exactly for lags  $\ell = 0$  through  $\ell = L$ . In practice,

this function must be determined from direct measurement on the data. Two common methods are currently being used. They are the Yule-Walker and Burg procedures. Yule-Walker (YW) involves the following three steps.

- 1) Estimate  $r_X(\ell)$  for  $\ell = 0, 1, \dots, L$
- 2) Substitute these values and solve for the AR coefficients  $a_1, \dots, a_L$ .
- 3) Compute a spectral estimate  $S_A(\omega)$  and the results of steps 1 and 2.

In Burg's method, the  $a_\ell$  coefficients and autocorrelation values are estimated simultaneously from the data values using a recursive algorithm. Once these parameters have been found, the spectral estimate  $S_A(\omega)$  is computed as in step 3 above. Burg's method will, in general, provide a different estimate than that given by the YW procedure. There is some evidence suggesting that higher resolution may be achieved for the case of sinusoids in white noise using Burg's algorithm, but this is still an open question for the case of random data.

The critical set of equations are the normal equations (to determine the coefficients) which frequent the literature. If the data is only available sequentially for example, it is possible to iteratively estimate the correlation coefficients and sequentially update them and in turn the coefficients. Similar work has been done in adaptive systems by Widrow et. al.

To briefly demonstrate the similarity, the MEM states that the least assumptions should be made about the unobserved points. This may be restated by saying that the spectrum estimated should be maximally random (maximum entropy) and still be consistent with the observed data. The equation  $\text{Entropy} = \int_0^\infty \log P(f) df$  (where intentionally different notation is used) is the entropy of a Gaussian stationary process. The object is to find a  $P(f)$  that maximizes this entropy and agrees with the measured values of the autocorrelation function. The MEM solution for this is

$$P(f) = \frac{P(N+1)/\omega}{\left| 1 + \sum_{n=1}^N \alpha(n+1) e^{-j2\pi f n \Delta t} \right|^2}$$

where  $P(N+1)$  and  $\alpha(n+1)$  are obtained from

$$\begin{bmatrix} c(0) & c(1) & \dots & c(N) \\ c(1) & c(0) & & \\ \vdots & \ddots & \ddots & \\ c(N) & \dots & c(0) \end{bmatrix} \begin{bmatrix} 1 \\ \alpha(2) \\ \alpha(3) \\ \vdots \\ \alpha(N+1) \end{bmatrix} = \begin{bmatrix} P(N+1) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This matrix equation may be recognized as being identical to that of the AR method as well as that resulting from designing a  $N+1$  point predictor.

#### Prony's Method

This is the oldest of the approaches considered with the original method being published by Prony in 1795 (26, 29), the same year as Gauss presented the theory of least squares estimation. It also has been rediscovered in recent years due mainly to the work of Van Blaricum, who will be presenting a review of this technique at this Workshop. A recent minisymposium dealing with the Prony approach (26) provides a good summary of the many facets of the method. Being well suited for modal analysis, it has gained popularity in several communities as a means for estimating the complex resonances of the system. These include the formulation of a voiced speech model for use in the linear prediction of speech (29), the characterization of numerical electromagnetic response data (26), etc. These resonances of a system are nicely represented in spectral data and may be considered the natural modes of the system. Whereas Prony's method is directed at finding the complex poles of the system which are comprised of both these natural frequencies and the damping factors, the spectral plot usually only shows the location of the natural frequencies which in many cases will be of sufficient use to the investigator. In this sense, Prony is more general. In summary, Prony is applied to estimate various types of spectral data particularly those where the interest is in finding major resonances.

The mathematical development of Prony has been presented many times and again appears in the development to be forthcoming by Van Blaricum. For completeness, this same development is repeated here.

The Prony approach is based on the well known fact that a system to be modeled can be represented by

$$R(t) = \sum_{i=1}^N A_i e^{s_i t}$$

where  $R(t)$  is the response, the  $s_i$  are the complex poles and the  $A_i$  are the corresponding residues. The  $s_i$  can be written as  $s_i = \alpha_i + j\omega_i$ . The  $\alpha_i$  are usually thought of as the damping constants, and the  $\omega_i$  are the natural frequencies in radians per second. In practice, the measured data usually appears as a set of discrete data; this equation may be rewritten as

$$R(t_n) = R_n = \sum_{i=1}^N A_i e^{s_i n \Delta t}, \quad n = 0, 1, \dots, M-1$$

where  $\Delta t$  is the time sample period and  $M$  is the total number of samples taken. This set of equations is  $M$  nonlinear equations in  $2N$  unknowns. If  $M$  is equal to or greater than  $2N$ , and if all  $\Delta t$  are equal, then this nonlinear set of equations can be solved using the Prony algorithm.

Prony further stipulated that the  $R_n$  must satisfy a difference equation of order  $N$  which may be written as

$$\sum_{p=0}^N \alpha_p R_{p+k} = 0, \quad k = 0, 1, \dots, N-1,$$

where  $N$  is the value of  $M-N$ . The roots  $z_i$  of the algebraic equation

$$\sum_{p=0}^N \alpha_p z^p = 0$$

define the natural frequencies through

$$z_i = e^{s_i \Delta t}, \quad i = 1, 2, \dots, N.$$

If the classic  $s$  to  $z$  transformation  $\alpha_N$  is defined equal to 1, then the remaining  $\alpha_p$ 's may be obtained by solving the equation

$$\sum_{p=0}^{N-1} \alpha_p R_{p+k} = -R_{N+k}.$$

If  $2N$  data samples are used, this last equation can be solved exactly for the  $\alpha$ 's, since the matrix equation has a unique solution (given certain conditions of independence). If more than  $2N$  samples are desired, then one can use a least-squares fit to obtain the solution since a unique solution is otherwise not existing. Once the  $\alpha_p$  have been found then the roots,  $z_i$  are found and the poles (major modes or resonances) are then obtained as

$$s_i = \frac{\ln z_i}{\Delta t}.$$

It is a simple procedure to obtain the residues,  $A_i$ , by solving the matrix equation once the  $s_i$  are known.

If pursued further, particularly in the case of noise where an exact fit is not possible, the mathematically inclined will recognize the similarities of this method with that of least squares, and the fitting of data using eigenvectors. In the latter, the residual error is equal to the sum of the remaining eigenvalues corresponding to the eigenvectors not used in the fitting of the data, a relation which is not uncommon in more detailed analysis of the Prony method.

In any case, this discussion should suffice as an introduction for the method, and the brunt of the explanation of its strengths and weaknesses will be left to the paper to be presented in the Workshop.

### Some Newer Techniques

In recent years there have been some innovative methods developed for spectral estimation. One of the most significant has been that by Papoulis (23) herein cited as the Papoulis algorithm. This method is applicable to bandlimited signals. The algorithm is an iteration involving the fast Fourier transform. In the reference, the convergence properties, the effects of noise, aliasing, etc. are described in detail. The method may be used to extrapolate bandlimited functions as well.

Given a bandlimited function  $f(t)$  and a segment of it  $g_0(t)$

$$G_0(\omega) = \int_{-T}^T g_0(t) e^{-j\omega t} dt$$

The  $n^{th}$  iteration step proceeds as follows: Form the function

$$F_n(\omega) = G_{n-1}(\omega) P_\sigma(\omega)$$

where

$$P_\sigma(\omega) = \begin{cases} 1 & |\omega| < \sigma \\ 0 & |\omega| > \sigma \end{cases}$$

by truncating  $G_{n-1}(\omega)$  and compute its inverse transform

$$f_n(t) = \int_{-\sigma}^{\sigma} F(\omega) e^{j\omega t} d\omega$$

Now form

$$g_n(t) = f_n(t) + [f(t) - f_n(t)] P_T(t) = \begin{cases} g(t) & |t| < T \\ f_n(t) & |t| > T \end{cases}$$

by replacing the segment of  $f_n(t)$  in the interval  $(-T, T)$  by the known segment  $g(t)$  of  $f(t)$ . Finally

$$G_n(\omega) = \int_{-\infty}^{\infty} g_n(t) e^{-j\omega t} dt$$

at the  $n^{th}$  step.

Note that  $f_n(t)$  is bandlimited and given by

$$f_n(t) = g_{n-1}(t) * \frac{\sin \sigma t}{\pi t}$$

By this continued resubstitution and multiple use of the FFT in transforming from one domain to the other, the bandlimited spectrum is estimated. In the time domain, since bandlimiting the spectrum is equivalent to

extending the signal (again the uncertainty principle), the extrapolation of the signal beyond the original duration interval provided may be performed as well. The procedure has been shown effective for bandlimited signals. A closed form procedure similar in approach, has been developed by J. Cadzow, where the numerical implementation does not require the truncation of generally infinite time signals, and therefore avoids the error producing truncations. More significantly, in the same work, Cadzow develops a closed form rule for generating the desired extrapolation in one step. W. Steenaart has also extended this class of extrapolation methods of Papoulis to a matrix formulation where the total process is achieved by one matrix operation resulting in savings in computation and yielding more accurate results in some cases. (These last two works are yet unpublished to my knowledge.)

Another new technique introduced by Gray is that of G-Spectral estimation. The transformation was introduced in 1971 and the application to spectral estimation was presented about 1976-1977 (31). It is especially valuable for processes whose autocorrelation can be expressed as a linear combination of complex exponentials. Results are still preliminary but tend to show a smaller mean squared error in some cases. A more extensive investigation is required to be able to critically compare the method.

There are many other methods as well. Woods (19) has extended his work to two dimensional processes and shows that his Markov model relates to the two dimensional maximum entropy spectrum. An iterative technique for computing an approximation to this spectrum is given as are results for real and simulated data. This Markov spectral estimate can offer higher resolution than other spectral estimates. A. K. Jain presents his work on two dimensional spectral estimation oriented to MEM in the Workshop, and will cover this area in more depth.

Much additional work has been done. For example, V. K. Jain has provided an analysis using a mathematical entity called a pencil of functions to be described later. Work by David Thomson on Spectral Estimation Techniques for Characterization and Development of WT4 Waveguide - I (BSTJ, Nov. '77) treats the case of spectral estimation for short and long records. Again, I will leave detailed discussion to Thomson in his paper in this Workshop also. Of course, not all work can be covered and my apologies to those omitted.

Overall, it is clear that there has been a major thrust in work relating to spectral estimation, using the revitalization of older methods with new twists, or by the development of newer methods. It is hoped that this Workshop, which additionally provides for the mechanism for comparison of these techniques, will answer some of the remaining questions.

#### Some Comparisons

The similarity between the AR and MEM approaches has been made several times so they are treated as essentially one technique at this point. Also the MLM and its relation to these techniques has also been brought out and is not separately discussed. The main thrust of the Workshop was motivated

by the AR and associated techniques as compared to the more conventional TT (and related methods) so this is the substance of the comparison.

It is fairly well accepted by many investigators that the AR methods exhibit higher resolution than the TT counterparts (Griffiths, Ulrych, Cooper). To be more exact, this statement must be modified to include the results that show this advantage to be true particularly for estimating a rational spectrum (one represented by a good all pole model). Not as great an advantage has been achieved with, for example, a Gaussian defined spectrum although based on results for sinusoids (rational) in noise, it is hoped that similar improvement can be made in improved resolution for non-rationally defined functions. In general, the AR methods will yield higher resolution for the same number of samples as the TT methods taking the above into account.

It is also concluded in many studies that the resolution is very much a function of the signal to noise ratio of the process being estimated. It has been shown (Marple) that the resolution is in fact variable and is a function of the signal to noise ratio. In contrast, the TT methods are more independent of noise. This major problem with AR and its sensitivity to noise has been to some extent reduced or traded-off by recent advances and improvements, but remains inherent with the approach and I view it as unavoidable. The characteristic of the approach is such as to estimate the coefficients from which the spectra is derived. These coefficients in actuality represent parameters in the denominator polynomial of the model used and naturally will be sensitive to errors (noise) as anything that leads to a matrix in the normal equations producing something short of a correlation matrix with a weak major diagonal. This phenomena is well known in system theory when similar difficulties are encountered in the estimation of poles for system identification, in LPE in statistics and pattern recognition to name a few. The AR methods as a result of the above perform well in estimating spectra with narrow peaks and in a high signal to noise environment (Kaveh). However, it is also well known that MEM techniques exhibit line splitting at lower S/N and frequency shifting at moderate S/N (Fougere).

Moreover, AR methods are particularly well suited for short record lengths. In fact it has been stated (Newman) that Berg is simply not practical for long records to go to the other extreme.

In addition to the classic advantages of AR methods for obtaining higher resolution on shorter records of data, and the difficulties with sensitivity to noise, there are other factors to consider not all of which can be presented here. For example, given a situation where there are two neighboring peaks of equal width in the spectrum to be estimated, the AR method will tend to emphasize the stronger of the two signals. A narrower variance of the estimate will be obtained and the results will be misleading in that they may indicate premature termination of the estimation procedure, or reflect on the signals as having different widths, strengths, etc. The AR methods are critically dependent on the number of terms used in the model, with too many or too few yielding erroneous results. Yet more work is required to determine a fast and sure way to arrive at the optimum model

representation and order (although the AIC approach has gone a long way in this direction).

There are of course advantages and disadvantages to be mentioned for techniques within this class of AR methods. For example, the MEM (or regressive) technique has been shown to be better than the MIM (Baggeroer), but develops a noisier estimate. These detailed intra class comparisons will not be elaborated further here since there are too many and for which in most part comments made rely on selected experimental results.

Since it is more familiar, the pros and cons of the TT methods are more well known. Suffice it to say that these approaches are long data oriented, and exhibit the classic trade-off between time and frequency resolution. If we have a smooth TT estimate and good resolution, the classic approach will generally have a better chance (Cooper).

Computationally, the AR and TT methods are comparable and have been shown to be so on many experimental results (Cooper, etc.).

Overall with respect to these approaches, one can present the two sides, the advantages and disadvantages, and the user based on his needs and evaluation of his problem, pays his money and takes his choice.

Comparisons can be made to some of the other methods as well which tend to bring out the similarities of these approaches by virtue of the similarities of difficulties that arise. For example, as it relates to Prony's method, questions which still need concrete answers include those involving the success of Prony's method with multiple poles, (addressed with some vigor by Van Blaricum) how to determine the order of the system apriori (this is exactly the same problem faced in AR with respect to the model), what are the effects of noise on Prony, are just a few (Van Blaricum - in a paper given in this Workshop). These are obviously not easy questions since they have been asked now for some 183 years when Prony's method was first published.

However, now comparisons are being made using results to be forthcoming in the symposium Workshop. Since I promised not to steal the thunder of the presenters to come (with many of them being able to exhibit more thunder than I can muster at this time), I will not say any more but simply address the coverage of each of the papers to be offered with an attempt to indicate where comparisons will be made along the way.

#### The Papers to be Presented

As of this writing, some 23 papers were scheduled to be presented in the Workshop. These are divided into four sessions, with some sessions dealing with more than one subject area. The organization of these papers as well as some selected comments follow.

The first three papers by Marple, Gersch, and Kaveh all deal with the autoregressive approach. Marple does a comparison of the conventional, autoregressive, and Pisarenko decomposition methods for the two sinusoid

case with results which tend to support those made in the comparisons of this paper. Gersch reviews the method and presents some detailed results and comparisons as well. Kaveh presents a rational spectral model and an efficient method for computing it and compares his result with the auto-regressive spectral estimator.

The next three papers by McAulay, R. deFigueiredo and Fougere deal with the maximum entropy or maximum likelihood approaches. McAulay, using state variables, compares the maximum likelihood estimator with the MEM, and relates the work to the Kalman filter in an excellent development. Fourgere describes one of the problems of the MEM, that of spontaneous line splitting in the presence of additive noise at low noise levels, and frequency shifts at moderate noise levels and offers some solutions.

The three papers that follow deal in one sense or another with extrapolative or recursive techniques, and include the works of Papoulis, Kailath, and Cadzow. Papoulis presents his work on adaptive extrapolation and hidden periodicities. Kailath discusses recursive spectral estimation; and Cadzow describes his work on incomplete data observations and extrapolation as well with applications to radar.

The next five papers broaden the work of the three major methods cited above with a wide variety of techniques. Starting out is Van Blaricum who presents a comprehensive review of Prony's method, the effects on noise on the method, and some examples. The use of pencil-of-functions approach is covered next by V. K. Jain et. al. with applications for system modeling and identification, another prospective to viewing the problem of spectral estimation. A. K. Jain then extends the work to two dimensional spectral estimation using maximum entropy with an iterative algorithmic approach. Lattice methods in spectral estimation is the topic discussed by Makhoul, and Thomson describes his work on non-parametric spectral estimates using Wishart distributions.

The remaining nine papers deal exclusively with applications of spectral estimation techniques. The first, by Tomlison et. al. is concerned with radar surveillance comparing MEM and LPC among others. Bowling's paper also deals with radar involving the MEM approach, for radar imaging. MEM and its application to Synthetic Aperture Radar (SAR) is handled by Jackson et. al., and gives some of the advantages and disadvantages of the technique for this problem. King and then Landers both consider MEM, the former applying it to wavenumber power spectra, and the latter to cepstral analysis. Two more radar oriented papers follow by Bowyer et. al., and Cooper and McGillem, dealing with radar clutter rejection and doppler spectral estimation respectively. The paper by Tsokos treats the case of spectral analysis of ionospheric data. Last but not least is the paper by Carmichael and Wiley who use classical zero crossing analysis to determine the results of one of the Workshop problems, and do very well at that.

This last paper is an excellent lead into the comparison of the methods on a common data set, the subject of the last day of the Workshop.

Part II of the Proceedings will be published in the form of a Technical Report and will compare the results of the different approaches as applied by the investigators to the three sample problems distributed as part of the Workshop. Together, these two reports should provide the reader with a comprehensive view of the field and means for evaluation, comparison and establishing directions for future research.

#### References

1. "Spectral Analysis - Theory, Implementation, and Applications," Rockland Systems Corporation.
2. M. Kaveh and G. Cooper, May 1976, "An Empirical Investigation of the Properties of the Autoregressive Spectral Estimator", IEEE Transactions on Information Theory, Vol. IT-22.
3. R. B. Blackman and J. W. Tukey, 1959, "The Measurement of Power Spectra from the Point of View of Communications Engineering."
4. E. Parzen, June 1968, "Statistical Spectral Analysis (single channel case) in 1968", Stanford University, Stanford, CA, Tech. Rep. 11 on Contract Nonr-225(80).
5. J. P. Burg, Oct. 1967, "Maximum Entropy Spectral Analysis", paper presented at 37th Annual International SEG Meeting, Oklahoma City, Okla.
6. A. B. Baggeroer, Sept. 1976, "Confidence Intervals for Regressive (MEM) Spectral Estimates", IEEE Trans. on Information Theory, Vol. IT-22, No. 5, pp. 534-545.
7. O. N. Strand, Aug. 1977, "Multichannel Complex Maximum Entropy (Autoregressive) Spectral Analysis", IEEE Trans. on Auto. Control, Vol. AC-22, No. 4, pp. 634-640.
8. J. Capon, Aug. 1969, "High-Resolution Frequency-Wave Number Spectrum Analysis", Proc. of IEEE, Vol. 57, pp. 1408-1418.
9. H. Akaike, June 1969, "Fitting Autoregressive Models for Prediction", Journal of Statistical Mathematics.
10. H. Akaike, Feb. 1970, "A Fundamental Relation Between Predictor Identification and Power Spectrum Estimation", Journal of Statistical Mathematics.
11. R. T. Lacoss, Aug. 1971, "Data Adaptive Spectral Analysis Methods", Geophysics, Vol. 36, No. 4, pp. 661-675.
12. T. J. Ulrych, March 1972, "Maximum Entropy Power Spectrum of Truncated Sinusoids", J. Geophysics Res., Vol. 77, pp. 1396-1400.

13. J. A. Edward and M. M. Fitelson, March 1973, "Notes on Maximum-Entropy Processing", IEEE Trans. Information Theory, Vol. IT-19, pp. 232-234.
14. W. Geisch and D. R. Sharpe, Aug. 1973, "Estimation of Power Spectra with Finite-Order Autoregressive Models", IEEE Trans. Auto. Control, pp. 367-369.
15. T. J. Ulrych, D. E. Smylie, O. G. Jensen, and G. K. C. Clarke, "Predictive Filtering and Smoothing of Short Records by Using Maximum Entropy", J. Geophysical Res., Vol. 78, No. 23, pp. 4959-4964.
16. H. R. Radoski, P. F. Fougere, E. J. Zawalick, "A High-Resolution Power Spectral Estimate: The Maximum Entropy Method", NTIS, AD-779 693 Tech. Report.
17. R. Jones, Dec. 1974, "Identification and Autoregressive Spectrum Estimation", IEEE Trans. on Auto. Control, Vol. AC-19, No. 6, pp. 894-898.
18. K. N. Berk, June 1973, "Consistent Autoregressive Spectral Estimates", Ann. Math. Stat., pp. 489-502.
19. J. Woods, Sept. 1976, "Two-Dimensional Markov Spectral Estimation", IEEE Trans. on Inform. Theory, pp. 552-559.
20. W. Newman, Jan. 1977, "Extension to the Maximum Entropy Method", IEEE Trans. on Inform. Theory, Vol. IT-23, No. 1, pp. 89-93.
21. J. M. Mendel, Oct. 1977, "White-Noise Estimators for Seismic Data Processing in Oil Exploration", IEEE Trans. on Auto.. Control, Vol. AC-22 No. 5, pp. 694-706.
22. F. J. Harris, Jan. 1978, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", Proc. of IEEE, Vol. 66, No. 1, pp. 51-83.
23. A. Papoulis, Sept. 1975, "A New Algorithm in Spectral Analysis and Band Limited Extrapolation", IEEE Transactions on Circuits and Systems.
24. L. J. Griffiths and R. Prieto-Diaz, Jan. 1977, "Spectral Analysis of Natural Seismic Events Using Autoregressive Techniques", IEEE Trans. on Geoscience Electronics.
25. L. Marple, "Resolutions of Conventional Fourier, Autoregressive and Special ARMA Methods of Spectral Analysis".
26. Mini Symposium on Modal Analysis of Experimental Data, Albuquerque, N.M., March 1977.
27. L. A. Gerhardt, May 1973, "Time Varying Spectral Disp ", AG'RD-NATO, Conference on Avionics, Rome, Italy.

28. L. A. Gerhardt, Nov. 1976, "Topologically Optimum Adaptive Networks", ICC, Washington, D.C.
29. J. Markel and Gray, 1976, "Linear Prediction of Speech", Springer Verlog.
30. A. Van den Bos, July 1971, "Alternative Interpretation of Maximum Entropy Spectral Analysis", IEEE Trans. on Info. Theory.
31. H. Gray, May 1977, "G-Spectral Estimation", IEEE Conf. on Acoustics, Speech and Signal Processing.

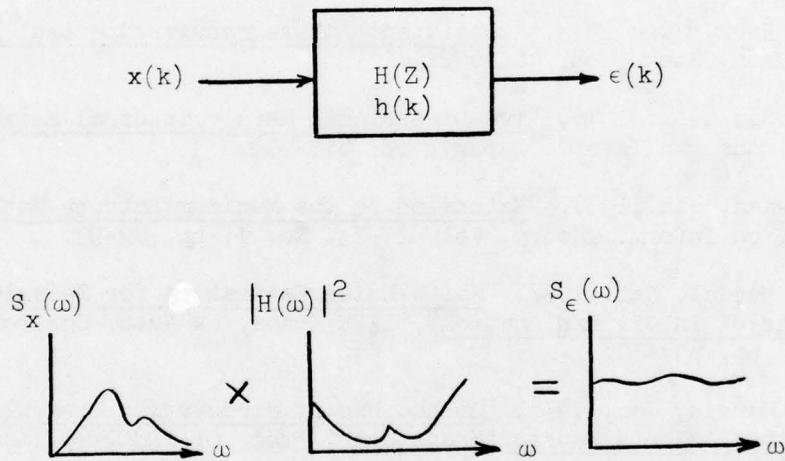


FIGURE 1. Whitening Filter Used to Obtain the AR Spectral Estimate

FREQUENCY RESOLUTION OF HIGH-RESOLUTION SPECTRUM  
ANALYSIS TECHNIQUES

LARRY MARPLE

Signal Science, Inc.  
3110 Coronado Drive  
Santa Clara, California 95051

Abstract

The two-sinusoid frequency resolutions of the conventional Fourier, autoregressive, and Pisarenko decompositon power spectral density estimates are presented. The conventional Fourier spectrum analysis method has a resolution which is, on the average, approximately the reciprocal of the observation (sample) interval. However, for any case of two sinusoids of some arbitrarily set initial phases, the resolution may be much greater or much less than the average resolution.

The Pisarenko decomposition method can, in theory, perfectly resolve two or more sinusoids if the number of sinusoids and the autocorrelation function is perfectly known. The autoregressive (maximum entropy) spectral estimation procedure, on the other hand, has a resolution that varies as a function of the signal-to-noise ratio (SNR). Its theoretical resolution ranges between that of the Pisarenko decomposition to that of the conventional Fourier methods. From experiments, it has been found that at a 20dB SNR, the autoregressive (AR) resolution performance was about four times that of the conventional Fourier. At 0dB SNR, the factor was about twice as good and at -10dB SNR, there was almost no difference in the resolution.

Introduction

A rule of thumb often stated when computing a power spectral estimate is that the frequency resolution is the reciprocal of the time interval from which data for the spectrum analysis has been collected. This rule is based on the uncertainty relationship between the time duration of a signal,  $\Delta T$ , and the "duration", or extent, of the signal transform,  $\Delta f$  [1]. Although the uncertainty relationship is normally concerned with one signal, it's interpretation as a resolution measure for two sinusoids assumes that the signals can be as close as  $\Delta f$  Hz apart before there is significant overlap of the transforms that will not permit the two separate responses of the two sinusoids to be distinguishable. Thus,

$$\Delta f \approx 1/\Delta T \quad (1)$$

Conventional analysis implicitly assumes that the signal outside the window of observation is zero. Any signal is arbitrarily truncated to have a duration of  $\Delta T_0$  seconds, the window width. For conventional methods, then,  $\Delta T = \Delta T_0$ . For the high resolution analysis methods, it will be shown that they have a non-zero extension to the observed data. The measured plus extended data then has an effective time duration  $\Delta T > \Delta T_0$ . As such, the frequency resolution of the high resolution methods will have better resolution than the conventional spectral methods. This is the basis for the high resolution claims of the auto-regressive (AR) methods. This paper quantifies both the theoretical and achievable resolutions with plots of the resolution performance for two conventional Fourier methods, the AR method using the Burg algorithm, and the Pisarenko decomposition procedure.

An illustration of the range of achieved resolutions using the same data for one conventional and two high-resolution techniques is shown in Figure 1. The AR-with-noise-power-cancellation (NPC) method is an approximation to the Pisarenko decomposition procedure. The details of the AR with NPC is discussed in reference [2].

#### Measure Of and Condition For Resolution

In this paper, the condition for two-sinusoid resolution is defined as the frequency  $\Delta f = |f_1 - f_2|$  at which the power spectral density (PSD) evaluated at the center frequency  $S(f_c)$ , where  $f_c = (f_1 + f_2)/2$ , is equal to the average of the PSDs evaluated at the two sinusoid frequencies, i.e.,

$$S(|f_1 + f_2|/2) = \frac{1}{2} [S(f_1) + S(f_2)] \quad . \quad (2)$$

This definition of resolution, shown in Figure 2, was motivated by the desire to provide a common method that could be applied to any method of PSD estimation since the Pisarenko decomposition and AR methods do not have the traditional mainlobe functions from which a 3 dB response width can be measured.

The measure of frequency resolution is given in terms of a dimensionless quantity  $R$  called the normalized resolution,

$$R = 2\pi M \Delta t \Delta f \quad (3)$$

where  $\Delta t$  is the sampling interval in seconds,  $\Delta f$  is the frequency separation in Hz at the point of just being resolved, and  $M$  is the number of autocorrelation lags. The motivation for the definition of the normalized resolution was taken from the time-bandwidth product relationship of the uncertainty principle. As such, for conventional spectrum analysis one would expect  $R \approx 2\pi$ , or about 6.28. To determine resolution in Hz, one simply looks up the proper  $R$  on a normalized resolution plot for given  $M$  and SNR. With  $R$ ,  $M$ , and  $\Delta t$  known, then

$$\Delta f = R / 2\pi M \Delta t \quad \text{Hz} \quad . \quad (4)$$

Note that  $M \Delta t$  is the total observation interval.

#### Conventional Fourier Methods Resolution

Based on the above definition of  $R$ , an analytically derived normalized resolution for the conventional Fourier method using known autocorrelation lags is shown on Figure 3. The details of the lengthy analysis is provided in reference [3]. The conventional Fourier method is the Blackman-Tukey PSD,

$$\hat{S}_{BT}(f) = \sum_{m=-(M-1)}^{M-1} \Phi_m \exp(-j2\pi f m \Delta t) \quad (5)$$

where the  $\Phi_m$  are the known or estimated autocorrelation lags. Note that the summation in equation (5) is finite, indicating the zero extension implied by the conventional method.

Often a weighting is used with the lags to reduce the effects of sidelobe leakage. A common weighting is the triangular, or Bartlett, window. When this window is used, the resolution degrades slightly to that shown in Figure 3. This illustrates how windowing will reduce sidelobes at the expense of resolution. Note that both the curves for the Blackman-

Tukey procedure fall below the 6.28 line given by the uncertainty principle. This indicates it is possible to achieve a resolution somewhat better than the standard rule of thumb.

The more common conventional Fourier PSD estimate is the periodogram, where the estimate is computed directly from the time samples  $x_m$ ,

$$\hat{S}_{per}(f) = \frac{1}{M\Delta t} \left| \sum_{m=0}^{M-1} x_m \exp(-j2\pi f_m \Delta t) \right|^2 . \quad (6)$$

This is most often computed with the FFT, which evaluates equation (6) at discrete intervals of the frequency parameter  $f$ . Figure 4 summarizes the normalized resolution for this method.

Figure 4 illustrates that in addition to the observation duration, the relative phasing between the two sinusoids is a large factor which determines the periodogram resolution. Since the periodogram is based on a windowed transform of the data, windowed sinusoids produce  $\sin f/f$  functions in the transform domain. Depending on the initial phases of the sinusoids, the net transform of two or more windowed sinusoids is the result of the complex vector constructive and destructive interference of the sidelobes of the  $\sin f/f$  functions. This sidelobe interaction as a function of initial phases of the two sinusoids has a great affect on the resolution achievable by the periodogram, as shown in Figure 4. Figure 4 was derived by running a computer program that determined the range of resolutions for successive  $5^\circ$  steps in the initial phases of the two sinusoids. No noise was added since the periodogram resolution is not a function of the SNR (however, the variance of the resolution over an ensemble of data sets with noise is a function of SNR). Therefore, the mean resolution points on Figure 4 are the average resolution performance for the periodogram over all possible phases. The average of the means gives a rough rule-of-thumb for the periodogram resolution of  $\Delta f \approx 0.86/\Delta T$  Hz, where  $\Delta T = M\Delta t$  is the observation interval. This contrasts with the usual  $1/\Delta T$  rule of thumb normally used.

The scalloped mean resolution curve has been reproduced on Figure 3. It overlaps the curve of the Blackman-Tukey PSD using known autocorrelations with Bartlett weighting. This is not surprising since the periodogram and Blackman-Tukey procedures yield identical results if the autocorrelation lag estimates

$$\hat{\Phi}_n = \frac{1}{M} \sum_{m=0}^{M-1-n} x_m x_{n+m}, \quad (7)$$

for  $n=0, \dots, M-1$  and letting  $\hat{\Phi}_{-n} = \hat{\Phi}_n$ , are used in equation (5). It is easily shown that equation (7) constitutes a biased estimate of the lags, the bias being that of a Bartlett weighting.

#### Pisarenko Decomposition Resolution

When a signal is known to consist of pure sinusoids in white noise, an appropriate procedure to find the unknown frequencies and powers of the sinusoids in the signal is the Pisarenko spectral decomposition procedure. In order to perfectly resolve the spectral components, i.e., decompose the spectrum perfect knowledge of  $M+1$  lags of the autocorrelation function for  $M/2$  sinusoids in white noise is required. Given these, Frost [4] has formulated a procedure based on Pisarenko's abstruse paper [5] to find the sinusoid frequency, summarized here.

- (A) Determine the smallest eigenvalue  $\lambda_{\min}$  of the autocorrelation matrix of rank  $M+1$ , where

$$\underline{\Phi}^{M+1} = \begin{bmatrix} \Phi_0 & \cdots & \Phi_M \\ \vdots & \ddots & \vdots \\ \Phi_M & \cdots & \Phi_0 \end{bmatrix} \quad (8)$$

The minimum eigenvalue is guaranteed to be the white noise power spectral density,  
 $\lambda_{\min} = \sigma_N^2$  (see Marple [3] for proof).

- (B) Solve for the eigenvector  $\underline{\chi} = \text{col}[1, \chi_1, \dots, \chi_M]$  corresponding to  $\lambda_{\min}$ ,

$$\underline{\Phi}^{M+1} \underline{\chi} = \sigma_n^2 \underline{\chi} \quad (9)$$

The vector  $\underline{\chi}$  may be related to the coefficients of an ARMA model of the sinusoids in white noise process [3].

- (C) The frequencies of the sinusoids are found by evaluating the roots of the polynomial

$$\underline{\chi}^T \underline{z} = 1 + \chi_1 z + \dots + \chi_M z^M = 0 \quad (10)$$

where  $\underline{z} = \text{col}[1, z, z^2, \dots, z^M]$ . The roots  $z_i$  are assured of having unit modulus, so that  $z_i = \exp(j\omega_i)$ , and  $\omega_i$  are the radian frequencies of the sinusoids.

A companion equation to the above procedure to obtain the power of each spectral component has been given by Marple [3, 4].

The perfect resolution of the Pisarenko decomposition (PD) procedure is depicted as the line of  $R=0$  in Figure 3. However, in practice one usually estimates the autocorrelation lags from data. Also, the number of sinusoid components is not typically known a priori. These two factors introduce error in the procedure of equations (8)-(10) and prevent the achievement of perfect resolution. A procedure known as AR with noise power cancellation is able to approach the potential performance of the PD without the need to solve a eigenproblem and a polynomial root problem. This method is discussed in another paper [2].

### AUTOREGRESSIVE METHOD RESOLUTION

Many claims have been made about the superior resolution of the autoregressive (AR) method, alias maximum entropy method (MEM). In the original MEM given by Burg [6], known auto-correlation lags were used to compute the PSD estimate given by

$$\hat{S}_{AR}(f) = \frac{P\Delta t}{\left| 1 + \sum_{m=1}^M a_m \exp(-j2\pi f_m \Delta t) \right|^2} . \quad (11)$$

The scalar and the AR coefficients  $\{a_m\}$  are found by solving the matrix equation

$$\underline{\Phi}^{M+1} \underline{A} = \underline{P} \quad (12)$$

where  $\underline{\Phi}^{M+1}$  is given by equation (8),  $\underline{A} = [1, a_1, \dots, a_M]^T$ ,  $\underline{P} = [p, 0, \dots, 0]^T$ .

It can be shown [3] that the AR PSD may be equivalently written as

$$\hat{S}_{AR}(f) = \sum_{n=-\infty}^{\infty} \phi_n \exp(-j2\pi f_n \Delta t) \quad (13)$$

where

$$\begin{aligned} \phi_n &= \phi_n && \text{for } |n| \leq M \\ \phi_n &= - \sum_{m=1}^M a_m \phi_{n-m} && \text{for } |n| > M . \end{aligned} \quad (14)$$

Thus, the AR PSD estimate has no implied window like the conventional Fourier methods (see equation 5) since the AR PSD is equivalent to a conventional method with infinite summation limits. No window means no sidelobe leakage phenomena with

the AR PSD. Also, the lag extension given by equation (14) gives the effective time duration  $\Delta T > \Delta T_o$ , meaning a higher resolution than the conventional PSD.

An analytical treatment of equation (11) using the auto-correlation function for two equal amplitude sinusoids in white noise provided the mean resolutions shown in Figure 3 as a function of SNR. For very high SNR, the performance approaches that of the PD procedure. For very low SNR, the performance asymptotically approaches that of the conventional Blackman-Tukey procedure.

Figure 5 replots the normalized resolution curves to a different scale format. An empirical equation which fits these composite curves is

$$R = 6.471 [SNR(M+1)]^{-.31} . \quad (15)$$

Note that the resolution is proportional to a power of the product of the SNR (in linear rather than dB units) and the number of lags.

When only data samples rather than the autocorrelation lags are available, the resolution performance is shown in Figure 6. The AR PSD is determined using the Burg algorithm [6] to determine the AR coefficients by a least squares estimation procedure. Many other algorithms are available to determine the AR coefficients. Makhoul [7] has formulated a lattice method algorithm. Morf et. al. [8] have formulated a ladder method algorithm. Time did not permit testing these techniques.

Two sinusoids were used, each of zero degrees initial phase, at three SNR levels. The zero phase condition was found to be about the worst case resolution. The periodogram resolution is shown for comparison. A large ensemble of data sets was run in order to determine the mean resolution and variance, as illustrated in Figure 6. From the figure, at 20 dB SNR the resolution of the AR method using the Burg algorithm is about four times better than that at the periodogram. At 0 dB SNR, the improvement factor is about two.

As the initial phases of the sinusoids change, the resolution plots of the AR and periodogram PSD techniques tend to track each other.

In the intermediate case, in which fewer AR coefficients than data points are computed, the performance is shown in Figure 7. As expected, an increase in the number of data points for fixed number of AR coefficients (4 in this case) brings the normalized resolution closer to the theoretical performance given in Figure 3.

The improvement in AR resolution illustrated thus far is perhaps optimistic. In practical situations, often no more than half as many AR coefficients as there are data points are calculated. This is done to minimize the extent of spurious responses in the AR spectra and to reduce the effect of spectral line splitting. Using fewer AR parameters than data values reduces the relative improvement of the AR spectrum over conventional spectra. As an example taken from Figure 3, consider the case when the SNR is -5 dB and 64 data samples are taken. The normalized resolution for the periodogram (FFT) for 64 points is about 5.4. The normalized resolution of the AR PSD for 32 coefficients and -5 dB is approximately 2.8. The relative resolution for this case is

$$\frac{\Delta f_{\text{ref}}}{\Delta f_{\text{AR}}} = \frac{R_{\text{ref}} / 2\pi 64\Delta t}{R_{\text{AR}} / 2\pi 32\Delta t} = \frac{5.4}{2(2.8)} = 0.96 \quad . \quad (16)$$

Since this ratio is less than one, the frequency resolution of the AR PSD estimate will, on the average, be worse than the resolution of a conventional FFT analysis. Thus, for low SNRs, there is normally no advantage to using an AR spectral estimate.

#### SUMMARY

The conventional Fourier methods have a mean resolution that is a function of the type and duration  $\Delta T$  of the window function. With the concept of resolution introduced in this paper, the frequency resolution  $\Delta f \approx 0.86/\Delta T$ . The AR method has a resolution that degrades to that of the conventional methods as the SNR decreases. Perfect resolution is obtainable with the PD method if the autocorrelation function is perfectly known or accurately estimated.

#### REFERENCES

1. Papoulis, Athanasios, 1977, Signal Analysis, McGraw-Hill Book Company, p. 273-278.
2. Marple, Larry, 1978, "High Resolution Autoregressive Spectrum Analysis Using Noise Power Cancellation", Record of the 1978 IEEE International Conference on Acoustics, Speech, and Signal Processing.
3. Marple, Larry, 1976, "Conventional Fourier, Autoregressive, and Special ARMA Methods of Spectrum Analysis", Engineers Dissertation, Department of Electrical Engineering, Standford University.
4. Frost, Otis L. III, 1976, "Power Spectrum Estimation", Proceedings of the 1976 NATO Advanced Study Institute on Signal Processing, Portovenere, Italy.
5. Pisarenko, V. F., 1973, "The Retrieval of Harmonics From A Covariance Function", Geophysical Journal of the Royal Astronomical Society, Vcl. 33, p. 347-366.
6. Burg, John P., 1975, "Maximum Entropy Spectral Analysis", PhD. Thesis, Department of Geophysics, Stanford University.
7. Makhoul, John, 1977, "Stable and Efficient Lattice Methods for Linear Prediction", IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-25, p. 423-428.
8. Morf, Martin, et. al., 1977, "Efficient Solution of Covariance Equations for Linear Prediction", IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-25, p. 429-433.

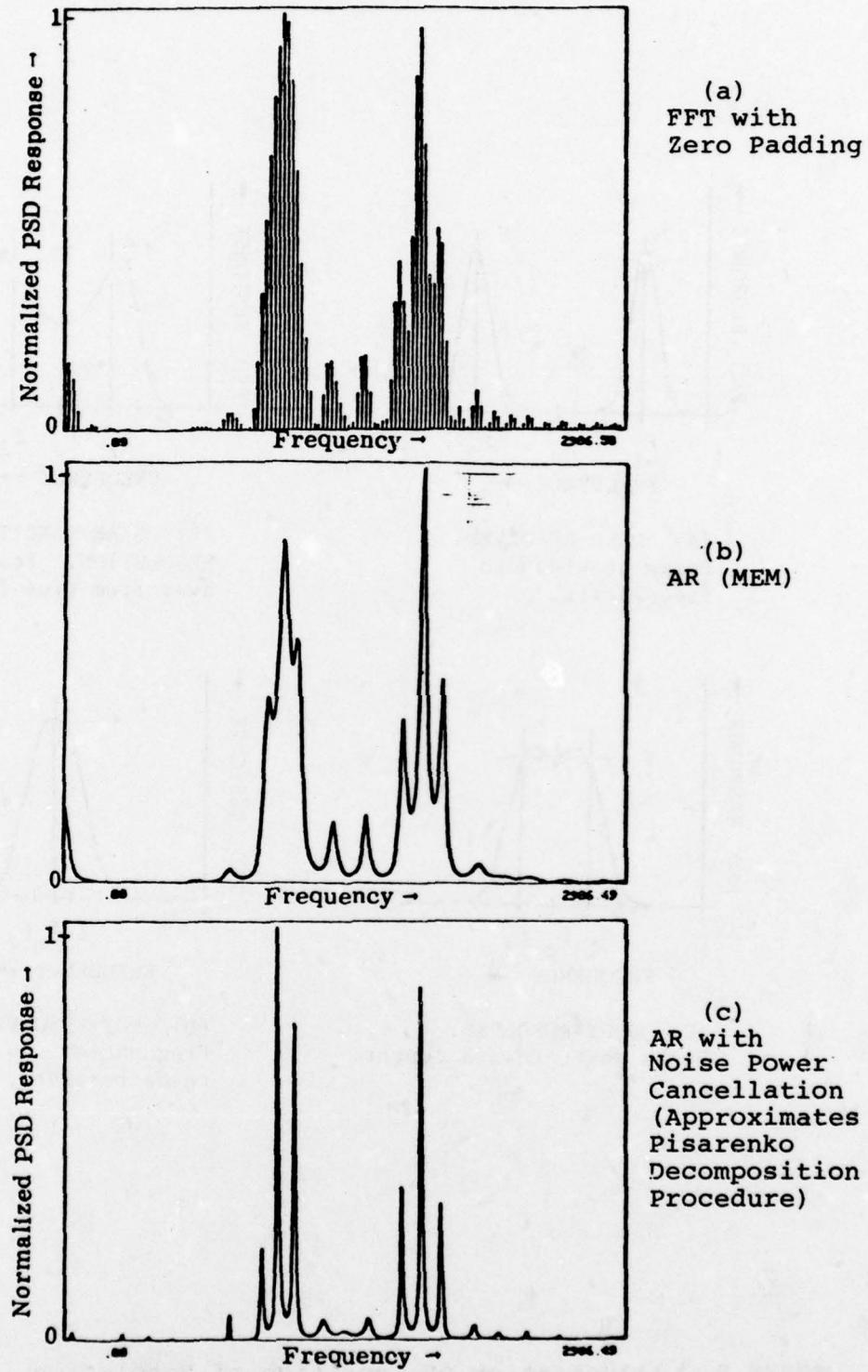
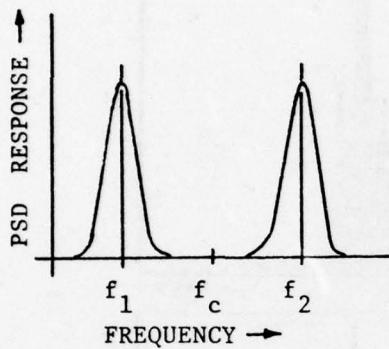
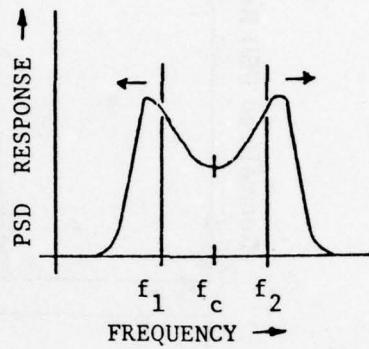


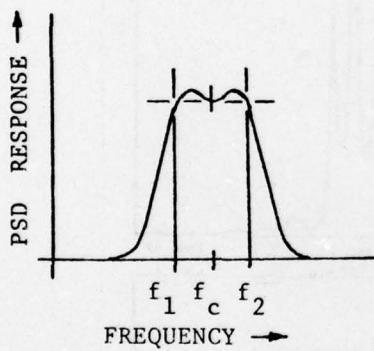
FIGURE 1. Illustration of Spectral Responses to a Sampled, Laboratory-Generated FSK Signal (25 dB SNR, 64 samples).



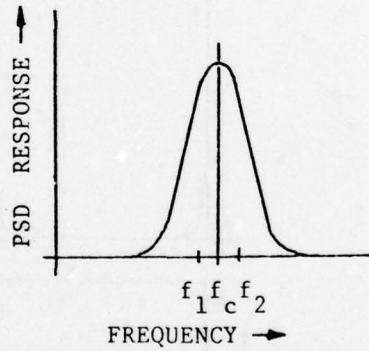
(A) WELL-RESOLVED.  
Peaks at sinusoid  
frequencies.



(B) NEAR CONDITION OF  
RESOLUTION. Peaks shift  
away from true frequencies.



(C) JUST-RESOLVED.  
Peaks shift toward center.



(D) NOT RESOLVED.  
Frequencies too close to  
to be resolved.

FIGURE 2. Illustration of Condition of Resolution.

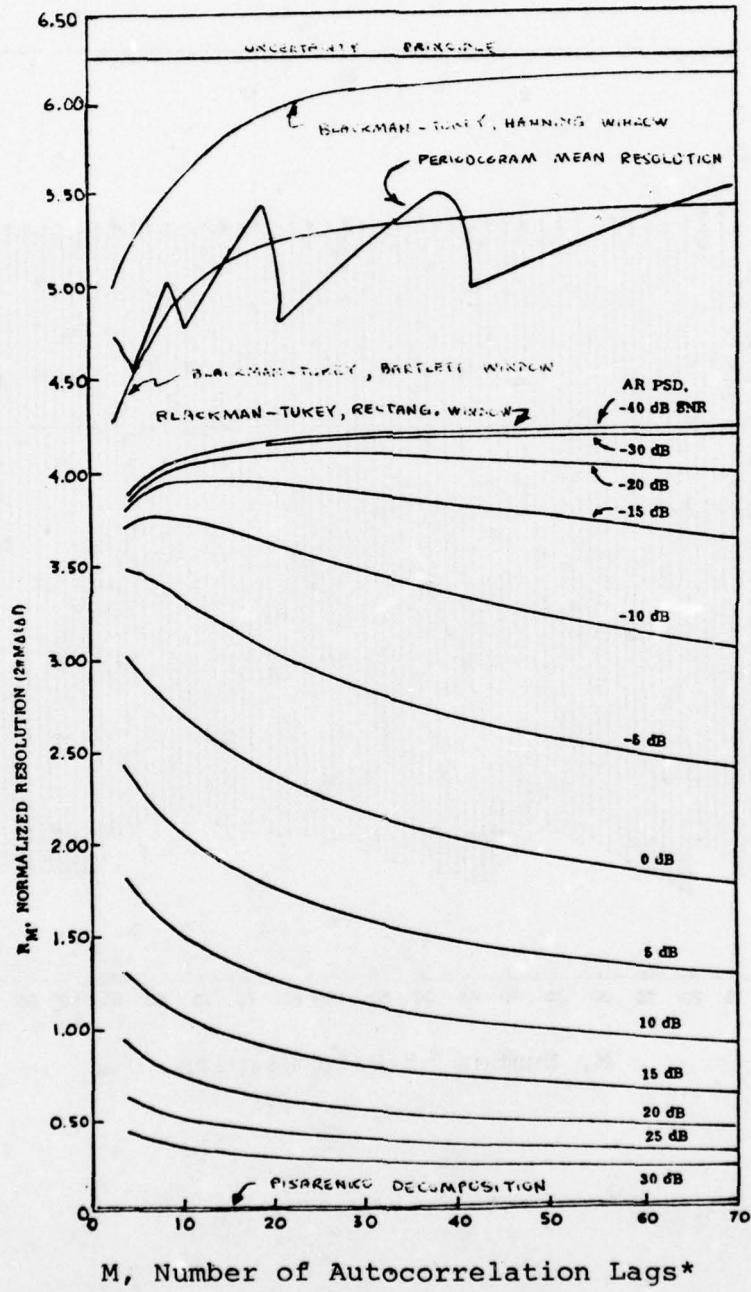
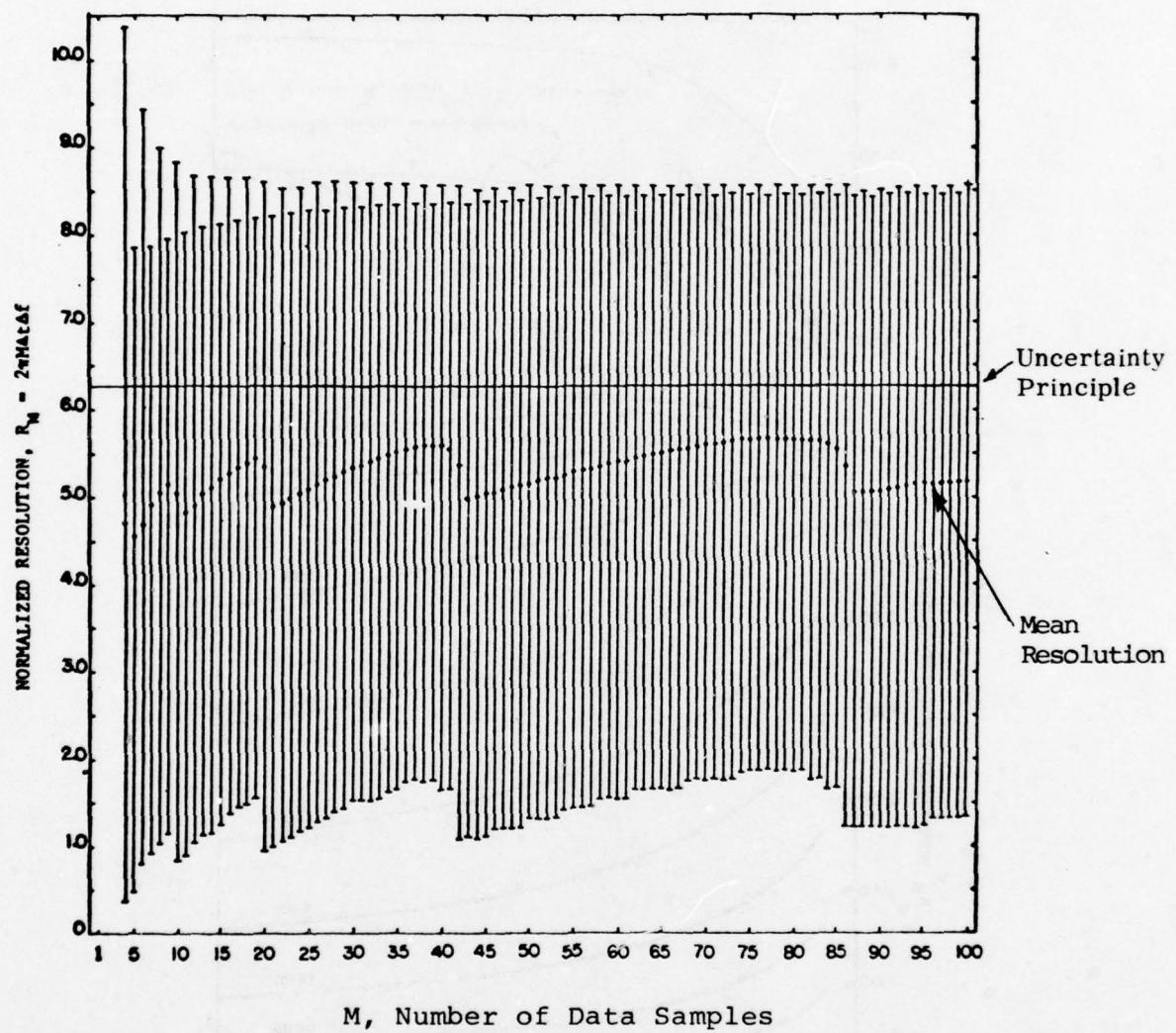


FIGURE 3. Normalized Resolution for Conventional, Autoregressive, and Pisarenko Methods

\*Number of AR Coefficients =  $M-1$



**FIGURE 4.** Periodogram (FFT) Resolution Variance As A Function of Initial Phase.

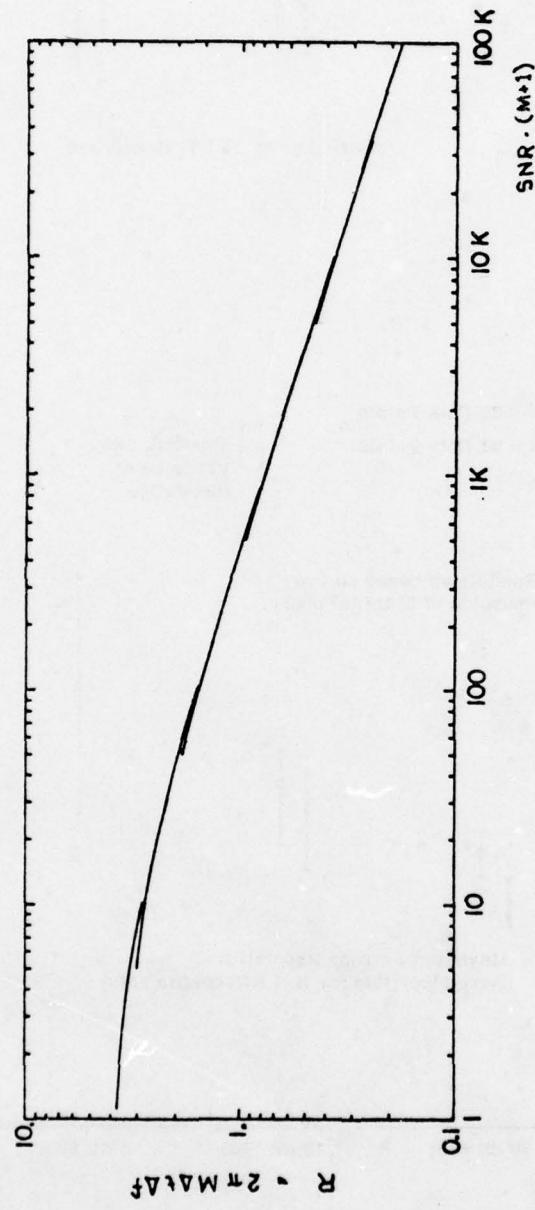


FIGURE 5. Normalized AR Resolution for Perfectly Known Autocorrelation Lags  
as a Function of  $\text{SNR} \cdot (M+1)$ .

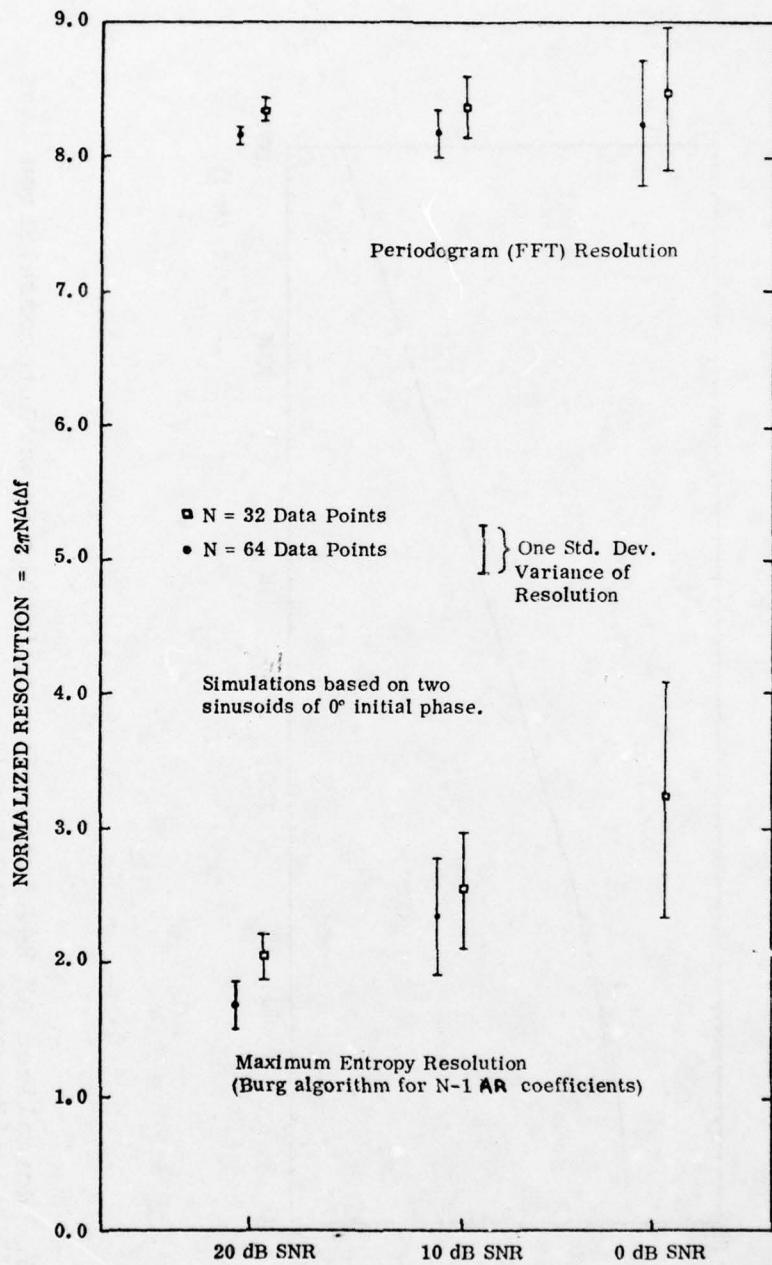


FIGURE 6. Normalized Resolution Using Data Samples Rather Than Known Autocorrelation Lags.

TWO-SIGNAL RESOLUTION, INTERMEDIATE CASE

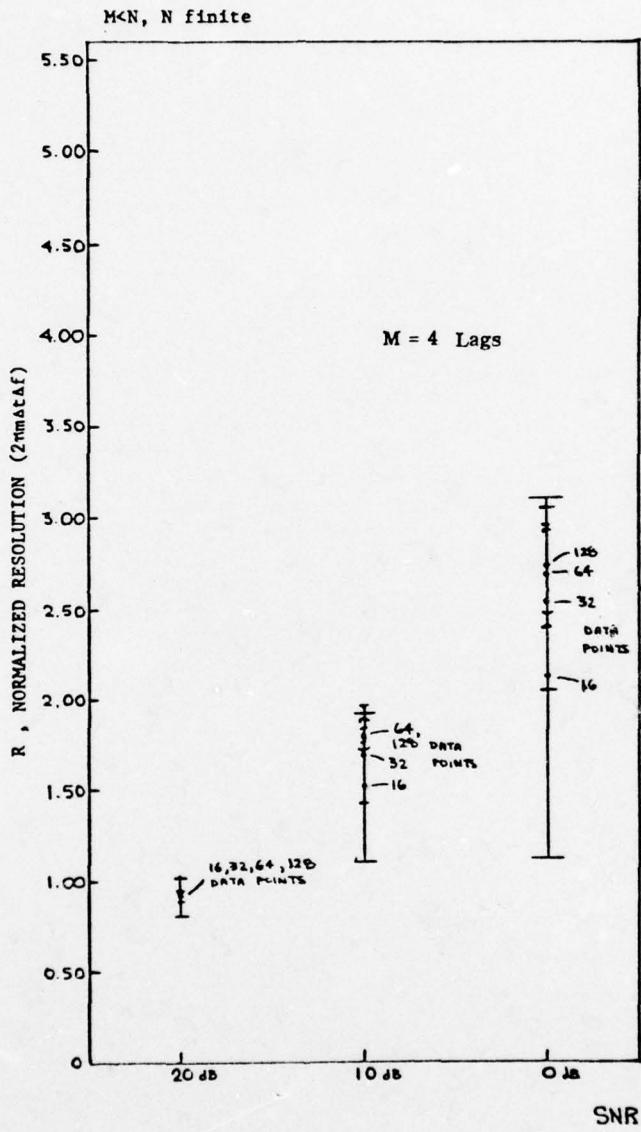


FIGURE 7. Normalized Resolution For An Intermediate Case In Which Fewer AR Coefficients Than Data Points Are Calculated.

AUTOREGRESSIVE MODEL SPECTRAL ESTIMATION,  
SOME SIMULATION STUDY STATISTICAL PERFORMANCE RESULTS.

WILL GERSCH and R.Z. LIU

Dept. of Information and Computer Sciences  
University of Hawaii  
Honolulu, Hawaii 96822

ABSTRACT

Some results of a simulation study of the statistical performance of autoregressive (AR) modeled spectral density and spectral coherence estimation are shown. The AR model order was determined by Akaike's AIC information theoretic criterion. Sharply peaked spectrum and smooth spectrum situations were considered. In the smooth spectrum situation, the statistical performance of the AR modeled spectral density and spectral coherence estimates in the vicinity of zero coherence, may be conservatively approximated by  $v = N/p_{\min}$  degrees of freedom where  $N$  is the number of data points and  $p_{\min}$  is the order of the AR model fitted to the data. Spectral estimation in the vicinity of sharp spectral peaks is poorest in the neighborhood of the peaks for AR modeled spectral estimation as it is for conventional windowed periodogram estimation. For sharp spectra, with large values of  $N$ , estimation of the spectral troughs and zero coherence appears to be well approximated by the  $v = N/p_{\min}$  property. Evidence for the asymptotic unbiasedness and consistency of AR modeled spectral estimation is also shown.

INTRODUCTION

There is an increasing interest in autoregressive (AR) parametric model methods for the spectral analysis of stationary time series data. The methods by which AR models are computed and the methods by which the AR model order is "best" fitted to the observed data are being actively researched. The statistical properties of spectral estimates obtained by AR models fitted to stationary time series data has had less attention. Only limited theoretical results of the statistical properties of AR modeled spectral estimates are known or conjectured, [1]-[6]. The status of the subject and our interest in particular applications in which AR modeled spectral analysis has an important role, [7]-[10], motivated a simulation study of some of the statistical properties of AR modeled spectral estimates in stationary vector time series.

We considered AR modeled spectral estimates by the Whittle-Akaike re-

cursive computation-AIC criterion model order method, [11]-[13], [9]. In that method, increasing order AR models are recursively fitted to the sample covariance data using Whittle's algorithm. The "best" AR model fitted to that data is that AR model whose order is determined by Akaike's AIC criterion. The AIC criterion selected AR model has an asymptotic minimum prediction variance property.

The phenomenology of AR spectral estimation is different than that of conventional windowed periodogram spectral estimation. Some of that phenomenology as well as some results on the asymptotic unbiasedness and consistency of the AR modeled spectral density and spectral coherence estimation are reported in this paper.

#### THE AR MODEL AND THE METHOD

$N$  consecutive samples of  $d$  simultaneous time series  $\{x(t); t=1, \dots, N\}$ , from an assumed covariance stationary time series are observed. The sample mean is deleted from each of the time series and the  $d \times d$  sample matrix covariance function

$$CXX(k) = \frac{1}{N} \sum_{t=1}^{N-k} x(t+k)x(t)' ; k=0, 1, \dots, L, \quad (1)$$

is computed from the remaining time series. In (1), and subsequently, ' denotes the matrix transpose operation.  $L$ , the maximum number of lags considered for analysis may be approximated by the empirically determined bound  $L \leq 3/\sqrt{N/d}$ .

An AR model of  $\{x(t)\}$  of order  $p$  is

$$x(t) = - \sum_{i=1}^p A^{(p)}(i)x(t-i) + \epsilon(t), \quad (2)$$

where the  $A(i)$  are  $d \times d$  matrices and  $(t)$  is a zero-mean  $d$ -vector innovations with covariance matrix  $V$ . The AR model coefficients satisfy the Yule-Walker equations,

$$\sum_{c=0}^p A^{(p)}(i)CXX(j-i) = 0; j = 1, \dots, p, A(0) = I \quad (3)$$

$$V_p = \sum_{c=0}^p A^{(p)}(i)CXX'(i).$$

The order of the "best" AR model fitted by Akaike's AIC criterion,  $p_{\min}$ , satisfies

$$p_{\min} = \min_p (N \log |V_p| + 2d(dp+1)), \quad p = 0, 1, \dots, L \quad (4)$$

The estimate of the  $d \times d$  spectral density matrix  $S(f)$  at frequency  $f$  is computed using the known formula, see [6] for example,

$$\begin{aligned} S(f) &= [I + A_1 e^{-j2\pi f} + \dots + A_{p_{\min}} e^{-j2\pi p_{\min} f}]^{-1} V_{p_{\min}} \\ &\quad [I + A'_1 e^{j2\pi f} + \dots + A'_{p_{\min}} e^{j2\pi p_{\min} f}]^{-1}. \end{aligned} \quad (5)$$

The square of the spectral coherence of frequency  $f$  between the time series  $x_i(t), x_j(t)$  is defined by,

$$W_{ij}^2(f) = \frac{S_{ij}(f) \cdot S_{ji}(f)}{S_{ii}(f)S_{jj}(f)} ; \quad i, j = 1, \dots, d, \quad i \neq j \quad (6)$$

where the average cross-power spectral density,  $S_{ij}(f)$ , is the  $ij$  element of  $S(f)$ .

Figure 1 shows AR modeled and Parzen windowed spectral analysis results for a  $d=2$  vector time series. The data analyzed corresponds to simulation of an AR-MA model of the random vibrations of a two degree-of-freedom structural system, [8], with  $N=1000$ . The AR model equivalent to a finite order AR-MA model is known to be of infinite order. The spectrum and coherences estimates achieved by the AR model fitted to the simulated data, (Fig(1)), are visually indistinguishable from the theoretical AR-MA model spectrum and coherence. The lag-50 Parzen windowed spectral analysis has reasonable coherence estimation performance but inadequate resolution. Also the Parzen window estimate of the major spectral peak is significantly under-biased. Increasing the spectral resolution by increasing the lag number would increase the variance of the estimates. That would introduce "bumpiness" in the spectrum removed from the peak and in the coherence. Thus, Figure 1 illustrates the known results that the spectral estimation of time series with sharp spectral peaks is difficult with conventional windowed periodogram methods and that parametric AR modeled spectral analysis can have relatively sharp resolution. These results motivated a statistical performance study in which AIC order determined AR modeled spectral analysis of a "worst case" sharp spectral peak time series example. A more familiar, "smooth" spectrum situation was also considered.

## RESULTS

### A: Sharp Spectral Peak Case

An AR-MA model with  $d=2$ , corresponding to the random vibrations of a two degree of freedom structure was used as the simulation model, [8]. Ten trials with  $N=500$ , and twelve simulation trials with  $N=1000$ , and  $N=2000$  data points were simulated. Table I shows the average and standard deviation of the AIC criterion determined AR model order,  $p_{min}$ , for the simulation trials.

TABLE I:  $p_{min}$  Average and  $p_{min}$  Standard Deviation for Different Simulation Data Lengths

	N=500	N=1000	N=2000
$p_{min}$ average	5.82	6.08	6.67
$p_{min}$ std. deviation	1.83	1.73	.65

As indicated in Table I and as expected, on the average, the length of the AR model, fitted to a sample function of a stationary time series of presumed infinite order, increases and the variability of the AR model order decreases with increasing data length  $N$ . For fixed  $N$ , the determinant of the matrix of the residuals,  $V_p$ , is a non-increasing function of  $p$ . In the vicinity of  $p=0$ ,  $V_p$  decreases sharply for increasing values of  $p$  and then tends toward a constant value, the innovations matrix variance, with increasing values of  $p$ . This effect plus the linear increase in the term  $2d(dp+1)$  in Eq(4) with  $p$ , account for the average increase of  $p_{min}$  with increasing  $N$ .

Figure 2a, b show the theoretical values and means, and Figure 2c, d the mean and mean  $\pm$  std. deviation of one of the two simulated time series AR modeled average power spectral densities for  $N=500$  and  $N=2000$ . These results are compatible with the theoretical results which suggest that the modeled average power spectral density tends to be unbiased and consistent.

In conventional windowed periodogram spectral analysis, the spectral estimate at frequency  $f$  is approximately distributed as a random variable which is a constant times a random variable which is distributed chi-square with degrees of freedom,  $\alpha X_v^2$ , where  $\alpha \approx S(f)/v$ ,  $v \approx 2S^2(f)/\text{var}S(f)$ , [14]. The only "known" related result on the behavior of AR modeled spectral estimates is conjecture by Parzen, [3], that AR modeled spectral estimates are asymptotically distributed in accordance with a complex Wishart distribution with  $v = N/p_{min}$  df.

Accordingly, to study the empirical statistical properties of AR spectral estimates, define  $v_T(f)$ ,  $v_B(f)$ ,  $v_V(f)$  respectively, the empirical total equivalent number of degrees of freedom, (df), the df due to bias and the df due to variance respectively at frequency  $f$ . These are given by

$$v_T(f) = \frac{2 \bar{\hat{S}}(f)}{\text{Var } \hat{S}(f) + (S(f) - \bar{\hat{S}}(f))^2}, \quad v_B(f) = \frac{2 S^2(f)}{(S(f) - \bar{\hat{S}}(f))^2},$$

$$v_V(f) = \frac{2 S^2(f)}{\text{Var } \hat{S}(f)}. \quad (7)$$

In Eq. (7),  $S(f)$ ,  $\hat{S}(f)$ ,  $\bar{\hat{S}}(f)$ , and  $\text{Var } S(f)$  are respectively the theoretical value, the estimated value, the mean and the variance of the spectral density at frequency  $f$ . Table 2 is a list of the quantities  $v_B$ ,  $v_V$  and  $v_T$  at the zero, first peak, trough, second peak and end point of the spectral density for  $N=500$  and  $N=2000$  for the time series modeled in Figure 2.

TABLE 2: Equivalent Degrees of Freedom of the Spectral Estimates

	N=500			N=2000		
	$v_B$	$v_V$	$v_T$	$v_B$	$v_V$	$v_T$
zero	749	73	67	8432	304	293
first peak	16	2	2	7945	17	17
trough	163	60	44	1568	585	425
second peak	29	55	19	2633	54	47
end point	346	43	41	14162	150	148

The following observations were abstracted from Table 2 and Figure 2:

- (i) For fixed  $N$ , df varies considerably as a function of frequency.
- (ii) For fixed  $N$ , df decreases as the sharpness of the spectral peaks increases.
- (iii) For fixed  $N$ ,  $v_T$  is larger at the troughs than at the peaks.

- (iv) For increasing  $N$ ,  $v_T$  may be approximated at the troughs by  $N/p_{\min}$   
 $(500/5.82 = 86, 2000/6.67 = 300)$
- (v) For increasing  $N$ ,  $v_T$  becomes dominated by the variance of the spectral estimates ( $v_V$ )

Result (i) is a "known" property of AR modeled spectral analysis. It holds for smooth as well as sharply peaked spectra. For smooth spectra, conventional spectral analysis yields relatively constant values of  $v$ . Results (ii) and (iii) are similar to those known for conventional spectral analysis. Result (iv) suggests that for sufficiently large  $N$ , the equivalent number of degrees of freedom,  $v_T$ , at the spectral troughs may be conservatively estimated by  $v \approx N/p_{\min}$ . Result (v) is consistent with Parzen's claim that AR spectral estimates tend to be (relatively) unbiased.

The sample properties of windowed periodogram spectral coherence estimates were treated by Akaike and Yamanouchi, [15], Jenkins and Watts [14], Brillinger [16] and Koopmans [17]. These results may be summarized by, [16],

$$\begin{aligned} E[W^2(f)] &\approx \hat{W}^2(f) + O(BW) + O(\frac{1}{BW \cdot N}) \\ \text{Var } \hat{W}^2(f) &\approx \frac{1}{v(f)} W^2(f) (1 - W^2(f))^2 + O(\frac{1}{BW^2 N}) \end{aligned} \quad (8)$$

where BW denotes "bandwidth" and  $\hat{W}^2(f)$  denotes the estimate of the spectral coherence at frequency  $f$ . From Eq. (8), it is seen that near zero coherence, the coherence estimates are dominated by bias errors and that as  $N$  increases, the variance of the conventional coherence estimate tends to zero and the bias tends to a constant  $O(BW)$ . Figure 3 shows the results for the theoretical, mean and mean  $\pm$  standard deviation of the spectral coherence for  $N=500$  and  $N=2000$  for the sharp spectrum example. The coherence estimates achieved by the AR model spectral estimates tend to be unbiased and consistent. The poorest statistical behavior is in the vicinity of  $W^2(f)$  equal to zero. Following Amos and Koopmans, [18], an exact expression for the distribution of  $z = \sqrt{W^2(f)}$  when  $W^2(f) = 0$  is

$$F(z) = 2(v - 1) z (1 - z^2)^{v-2}.$$

Using the integral expression for the beta function, the mean and variance of  $\hat{W}^2(f)$  when  $W^2(f)=0$  becomes

$$\begin{aligned} E[\hat{W}^2(f)|W^2(f)=0] &= \frac{1}{v} \\ \text{Var}[\hat{W}^2(f)|W^2(f)=0] &\approx \frac{1}{v^2} (1 - \frac{2}{v}) \end{aligned}$$

Using the results of Eq. (10) as a guide, an approximate equivalent number of degrees of freedom was computed for the estimate of zero coherence for the AR modeled spectral estimates based on the bias and variance of the coherence estimates in the vicinity of  $f=.2$ . ( $\hat{v}_B(f) = (\hat{W}^2(f))^{-1}$ ;  $v_Y(f) \approx (\text{Var } \hat{W}^2(f))^{-2}$ ). These results, with the Parzen conjectured df, are in Table 3.

TABLE 3: Degrees of Freedom for Zero Coherence Estimates

	N=500	N=2000
$\hat{v}_B$	23	294
$\hat{v}_Y$	25	243
N/pmin	86	300

The data in Table 3 indicates that the df estimated by the bias and the variance of the zero coherence estimates are very similar. This evidence supports the notion of an equivalent number of degrees of freedom for the estimation of zero coherence.

#### B: Smooth Spectral Density Case

For this situation an AR model of  $p_{\min}=7$  fitted to an electroencephalogram (EEG) time series with  $d=3$  was selected as the simulation model. Ten realizations each for data lengths  $N=200$  and  $N=800$  were simulated. Figures 4a, 4b show the mean spectral density and the mean  $\pm$  standard deviation of the spectral density computed from the synthesized data respectively for  $N=200$  and  $N=800$  for one time series. Figures 4c, 4d show the mean coherence and mean  $\pm$  standard deviation of the coherence computed for two data channels for  $N=200$  and  $N=800$  respectively. The true model values of the spectral density and spectral coherence are extremely close to the mean values for  $N=800$  in Figures 4b, 4d.

The illustrations of the spectral density and spectral coherence for the smooth spectral density estimates also suggest that AR spectral estimates are asymptotically unbiased and consistent. In contrast with the sharply peaked spectrum situation, in the smooth spectrum situation, the largest spectral deviations occur at the troughs of the spectrum. Using the notion of the df

due to variance,  $v_T$ , defined in Eq. (7), the df for N=200 and 800 respectively were 30 and 150. Correspondingly, the Parzen df estimates are 29 and 114. Using the definition of df for coherence estimates, Eq. (8), we obtained  $v_{200} \geq 30$ ,  $v_{800} \geq 120$  in the vicinity of the estimate of zero coherence. Thus  $v \approx N/p_{\min}$  appears to be a reasonably conservative estimate to use in evaluating smooth spectrum AR spectral estimate performance.

#### SUMMARY

Classical windowed periodogram spectral analysis is an empirical procedure. In that procedure, "window carpentry" is subjectively performed to balance bias and variance errors. The empirical-subjective nature of spectral analysis is generally ignored and frequently users of spectral analysis subroutines do not have sufficient expertise to properly interpret the results of their computations. In contrast with the classical method of spectral analysis, by the use of Whittle's recursive computational procedure and Akaike's AIC criterion, autoregressive model spectral estimates can be computed by an automatic computational procedure that has nice statistical properties.

Recent theoretical results on AR modeled spectral analysis estimates suggest that they are asymptotically unbiased and consistent and that their variance is as low as that achieved by the best windowed periodogram spectral estimators. The empirical evidence obtained from simulation trials shown here suggest that Akaike AIC criterion-AR model spectral estimates are asymptotically unbiased and consistent. The evidence also supports Parzen's conjecture that asymptotically, for smooth spectrum situations, the number of degrees of freedom in a spectrum estimate or in the estimate of zero coherence can be conservatively estimated as  $v=N/p_{\min}$ . Spectral estimates in the vicinity of sharp spectral peaks are likely to have large bias and variance errors. Asymptotically the estimates of the troughs of sharply peaked spectral time series do appear to follow the  $v=N/p_{\min}$  law. AR modeled spectral estimates do clearly have the resolution property to reproduce sharp spectral peaks without introducing the bumpiness or large variance that is characteristic of conventional windowed periodogram spectral analysis.

#### REFERENCES

1. Akaike, H., 1969, "Power Spectrum Estimation Through Autoregressive Model Fitting", Ann. Inst. Stat. Math.
2. Kromer, R. E., 1969, "Asymptotic Properties of the Autoregressive Spectral Estimator", Tech. Rept. No. 13, Dept. of Statistics, Ph.D. Dissertation--Stanford University.

3. Parzen, E., 1970, "Multivariate Time Series Modeling, in Multivariate Analysis II", P. R. Krishnaiah, Ed., Academic Press, N.Y.
4. Berk, K. N., 1974, "Consistent Autoregressive Spectral Estimate", Annals of Stat. 2, P. 489-502.
5. Huzii, M., February 1976, "On a Spectral Estimate Obtained by Fitting an Autoregressive Model Fitting", Technical Report No. 22, Stanford Dept. of Statistics, T. W. Anderson, Project Director.
6. Hannan, E. J., 1970, "Multiple Time Series", Wiley, N.Y.
7. Gersch, W. and Tharp, B. R., 1976, "Spectral Regression - Amount of Information Analysis of Seizures in Humans, in Quantitative Analytic Studies in Epilepsy", Kellaway, P. and Petersen, I. S., Edits., Raven Press, New York, P. 508-532.
8. Gersch, W. and Yonemoto, J., 1977, "Synthesis of Multivariate Random Vibration Systems: A Two Stage Least Squares AR-MA Model Approach", Jour. of Sound and Vibration, 52(4) P. 553-565.
9. Gersch, W. and Yonemoto, J., 1977, "Parametric Time Series Models for Multivariate EEG Analysis", Computers and Biomedical Research, 10, P. 113-125.
10. Gersch, W. and Yonemoto, J., 1977, "Automatic Classification of Multivariate EEGs Using an Amount of Information Measure and the Eigenvalues of Parametric Time Series Model Features", Computers and Biomedical Research, 10, P. 297-318.
11. Whittle, P., 1963, "On the Fitting of Multivariate Autoregressions and the Approximate Canonical Factorization of a Spectral Density Matrix", Biometrika 50, P. 129-134.
12. Akaike, H., 1974, "A New Look at Statistical Model Identification", IEEE Trans. on Automatic Control AC-19, P. 716-723.
13. Akaike, H., 1976, "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion", System Identification: Advances and Case Studies, Mehra, R. K. and Lainotis, P. G., Edits., Academic Press, P. 27-96.
14. Jenkins, G. M. and Watts, D. G., 1968, "Spectral Analysis and Its Applications", Holden-Day, San Francisco.
15. Akaike H. and Yamanouchi, Y., 1962, "On the Statistical Estimation of Frequency Response Function", Ann. Inst. Statist. Math., P. 23-56.

16. Brillinger, D. R., 1975, "Time Series, Data Analysis and Theory", Holden-Day, San Francisco.
17. Koopmans, L. H., 1974, "The Spectral Analysis of Time Series", Academic Press, N.Y.
18. Amos, D. E. and Koopmans, L. H., 1963, "Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes", Sandia Corp. Monograph SCR-483, Albuquerque, New Mexico.

### FIGURES

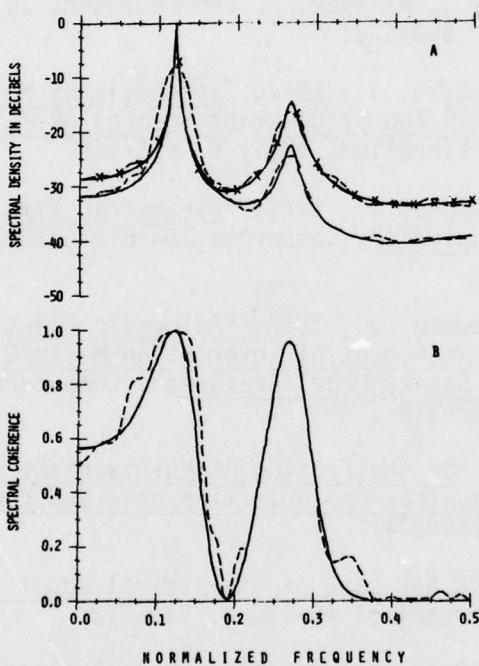


FIGURE 1: AR modeled pmin 7 (solid lines) and Parzen lag 50 windowed periodogram spectral analysis (dashed lines) of simulated structural system random vibrations.

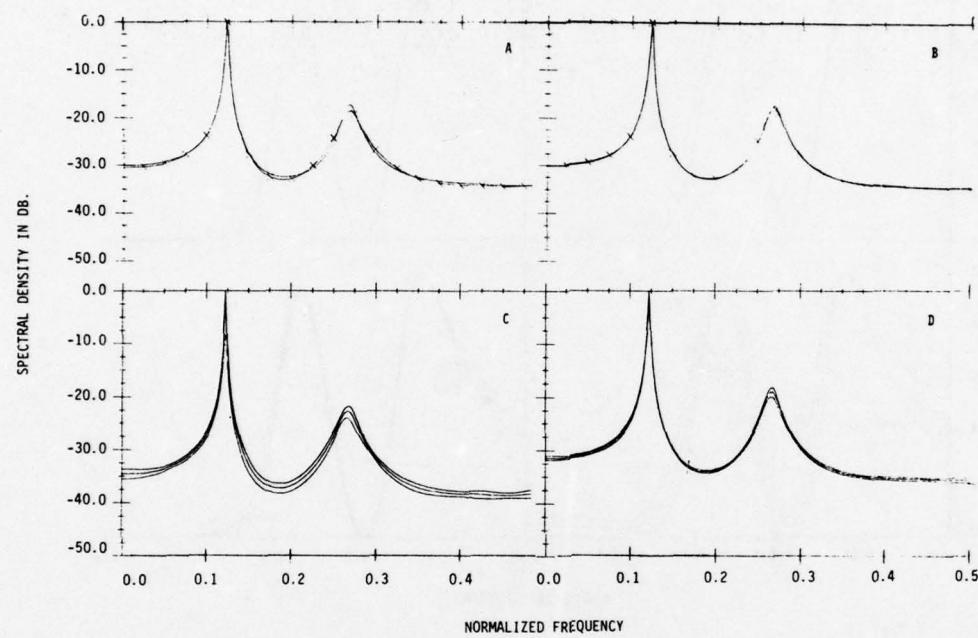


FIGURE 2: Theoretical and mean values (a)  $N=500$ , (b)  $N=2000$ ; mean and mean  $\pm$  standard deviation, (c)  $N=500$ , (d)  $N=2000$  of the average power spectral density versus frequency.

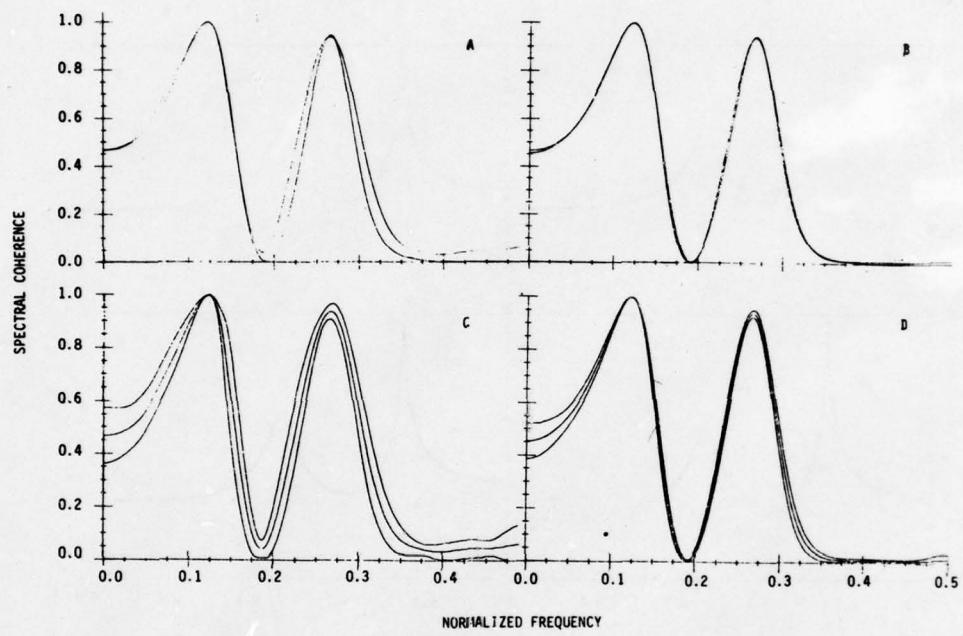


FIGURE 3: Theoretical and mean values (a)  $N=500$ , (b)  $N=2000$ : mean and mean  $\pm$  standard deviation, (c)  $N=500$ , (d)  $N=2000$  of the spectral coherence squared versus frequency.

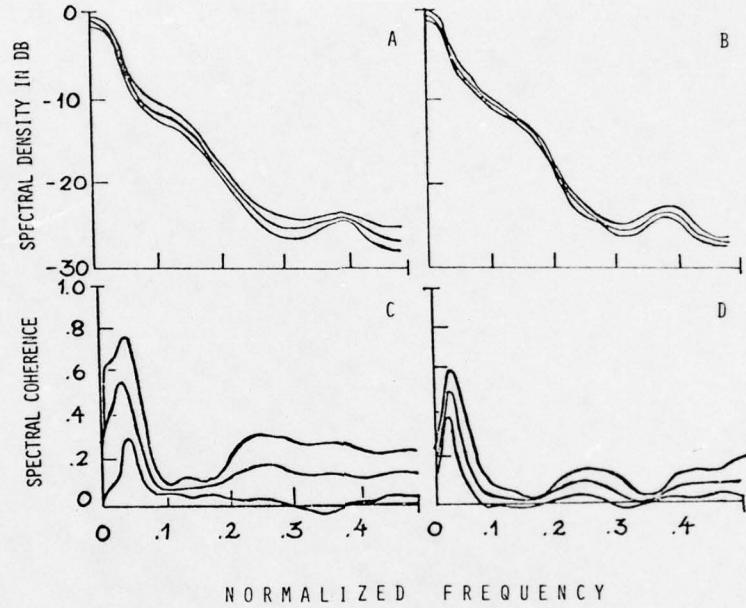


FIGURE 4: Mean  $\pm$  standard deviation; spectral density (a)  $N=200$ ,  
(b)  $N=800$ ; spectral coherence squared, (c)  $N=200$ , (d)  $N=800$ .

# HIGH RESOLUTION SPECTRAL ESTIMATION VIA RATIONAL MODELS

MOSTAFA KAVEH

Department of Electrical Engineering  
University of Minnesota  
Minneapolis, Minnesota 55455

## Abstract

A rational spectral model is derived for high resolution spectral estimation of data containing autoregressive signal, interference and white noise.

A computationally efficient method for estimating the rational spectrum is introduced. Examples of the spectra calculated by this method are presented and compared with autoregressive spectral estimator.

## Introduction

Recent interest in power spectral estimation has been focussed on data adaptive methods. This is because such methods are free from the effects of fixed window functions associated with the traditional Blackman and Tukey type methods for the estimation of spectra. One class of data adaptive spectral estimators is based on rational spectral models. Such a method was proposed by Tretter and Steiglitz [1] among others e.g. [2]. The general rational model for spectra, however, has not gained popularity in practice due to computational complexities and lack of understanding of the statistical properties of such estimators.

A subclass of rational spectral models, however, has become very popular in high resolution spectral estimation applications. This subclass is the all-pole model and includes the familiar autoregressive (AR) and maximum entropy (MEM) spectral estimators.

The properties of the AR spectral estimator have been studied, theoretically in the asymptotic case [3], [4], [5] and empirically [6] - [9]. It has been shown that this estimator in many cases offers considerably higher resolution based on the same amount of data, than the Blackman and Tukey type estimators. Furthermore, the above asymptotic and empirical investigations have shown the variance of the AR estimates to be comparable to the unsmoothed Blackman and Tukey type estimates, for the same number of autocorrelation lags. It should be pointed out, however, that the AR estimates usually require much fewer lags for the same resolution thus avoiding the problem of instability in the estimates.

One problem that has usually not been discussed in conjunction with the AR estimator has been its behavior in resolving spectral peaks from noisy data. Whereas theoretical treatment of this problem is difficult, it is one of great practical importance in spectral estimation where the desired signals are usually buried in noise. Two approximate and empirical studies on the resolution of AR spectral estimators have pointed out the dependence of its resolution on the signal-to-noise ratio. Lacoss [6] and Kaveh and Cooper [8] have shown that the order of the AR spectral estimator for the same resolution of a sinusoid and that of a narrow-band process with additive wide-band independent noise, to be approximately equal to the square root of the signal-to-noise ratio. Thus, for noisy signals one may have to resort to a high order estimator for its resolving capabilities and, therefore, become vulnerable to unstable estimates. To remedy this situation, this paper discusses spectral estimation with models containing both poles and zeros. Specifically, rational models are introduced as a means of resolving the spectra of narrowband signals, with  $L^{\text{th}}$  order AR models, in the presence of noise and interference. Preliminary examples using an ad-hoc method for the estimation of spectra are presented.

## II. The Autoregressive (AR) Spectral Estimator

A zero-mean time series  $\{x_t\}$  is said to satisfy an  $L^{\text{th}}$  order autoregressive model if:

$$x_t = \sum_{i=1}^L \alpha_i x_{t-i} + u_t \quad (1)$$

where  $\{\alpha_i\}$  denote the AR coefficients and  $\{u_t\}$  is a zero-mean uncorrelated (white) sequence. Another interpretation of the model in (1) is that  $\{\alpha_i\}$  represent an  $L^{\text{th}}$  order one-step ahead linear predictor of  $\{x_t\}$ . If  $\{\alpha_i\}$  are then estimated, based on a minimum mean square error criterion  $\{u_t\}$  on the average becomes an orthogonal sequence. It can be shown [10] that the model in (1) leads to a spectral density of the form

$$S_L(f) = \frac{S_1}{\left| 1 - \sum_{k=1}^L \alpha_k e^{i2\pi kf\Delta T} \right|^2}, \quad |f| \leq \frac{1}{2\Delta T} \quad (2)$$

where  $S_1$  is the spectral level of the  $\{u_t\}$  sequence.

The spectral density shown in (2) is the AR spectrum and it is this model that is fitted to an observed time series, by simply estimating  $\{\alpha_i\}$  from the time series. Several estimation procedures for  $\{\alpha_i\}$  have been discussed in the literature such as the maximum likelihood, the least-squares

[10] and a method based on forward and backward prediction error filtering [11]. If the number of data samples is not very small, the simplest and computationally most efficient estimates of  $\{\alpha_i\}$  are from a solution of the Yule-Walker equations. These equations arise, simply by multiplying equation (1) by  $x_{t-i}$ ,  $i=1, \dots, L$  and taking the expectation of both sides, to obtain a relation between the autocorrelation function of the process  $\{x_t\}$  and the coefficients  $\{\alpha_i\}$ . The Yule-Walker equations are then given by:

$$R_o A = \rho_o \quad (3)$$

where

$$R_o = \begin{vmatrix} r_0 & r_1 & \dots & r_{L-1} \\ r_1 & & & \\ \vdots & & & \\ r_{L-1} & & r_0 & \end{vmatrix}, \quad A = \begin{vmatrix} \alpha_1 \\ \vdots \\ \alpha_L \end{vmatrix} \quad \text{and} \quad \rho_o = \begin{vmatrix} r_1 \\ \vdots \\ r_L \end{vmatrix}$$

and where  $r_i$  is the autocorrelation function of  $\{x_t\}$  at the  $i^{\text{th}}$  lag. Furthermore, the power in  $\{u_t\}$  can be found by multiplying (1) by  $x_t$  and taking the expectation as:

$$S_1 = [r_0 - \sum_{i=1}^L \alpha_i r_i] \Delta T, \quad \Delta T \text{ the sampling interval} \quad (4)$$

In practice  $\{\alpha_i\}$  and  $S_1$  are estimated from (3) and (4) based on estimates of the autocorrelation function  $\{\hat{r}_i\}$ . Figure [1] shows a comparison of the spectral estimates of a radar doppler spectrum by the AR method and the traditional Blackman and Tukey method with a Hanning taper. The resolving capability of the AR method based on much fewer lags  $L$  is quite evident in this example.

### III. Spectral Estimation in the Presence of Noise and Interference

We now assume that the signal  $\{x_t\}$  satisfies an  $L^{\text{th}}$  order AR model and therefore its spectrum can be estimated as in the previous section. The problem of interest is the estimation of the spectrum of the observed signal

$$y_t = x_t + \omega_t + n_t \quad (5)$$

where  $\{\omega_t\}$  is an AR( $M$ ) process and considered to be the interference and  $\{n_t\}$  is a white noise sequence, with  $\{n_t\}$ ,  $\{x_t\}$  and  $\{\omega_t\}$  mutually uncorrelated.

The resolution of the spectral estimators are now discussed in the asymptotic case, that is, when the autocorrelation function of  $\{y_t\}$  is accurately known.

Let  $\{x_t\}$  be described by (1) and  $\{\omega_t\}$  be given by the following autoregressive model

$$\omega_t = \sum_{i=1}^M b_i \omega_{t-i} + v_t$$

The z-spectrum of  $y_t$  is then given by:

$$S_Y(z) = \frac{S_1}{D_x(z)D_x(z^{-1})} + \frac{S_2}{D_\omega(z)D_\omega(z^{-1})} + S_n \quad (6)$$

where  $S_1$  and  $S_2$  are given by (4) using the appropriate autocorrelation values for  $\{x_t\}$  and  $\{\omega_t\}$ ,  $S_n$  is the spectrum of  $n_t$  and

$$D_x(z) = 1 - \sum_{i=1}^L a_i z^i, \quad D_\omega(z) = 1 - \sum_{i=1}^M b_i z^i$$

Putting (6) under a common denominator, it becomes obvious that  $S_Y(z)$  is the spectrum of an autoregressive moving average process of orders  $L+M$  and  $L+M$ , i.e., AR( $L+M$ )/MA( $L+M$ ). This is a process with  $L+M$  zeros and  $L+M$  poles, where, from equation (6), the numerator polynomial coefficients are related to  $\{a_i\}$ ,  $\{b_i\}$  and  $S_n$  in an obvious manner.

It can be seen that the estimation of  $S_Y(z)$  using a purely AR technique (all-pole) in equivalent to approximating the ( $L+M$ ) order moving average component by an AR one. This, theoretically would require an infinite order model. Finite order models of order  $L$ , however, will estimate  $S_Y(f)$  with good resolution, with  $L$  depending on the various parameters of signal noise and interference, notably their relative power.

It is obvious now that whereas the resolution of Blackman and Tukey type spectral estimations are only determined by the window bandwidths in a predictable fashion, those of the AR and generally ARMA estimators are very much data dependent, requiring larger all-pole orders or ARMA modeling.

The general problem of the identification of ARMA models for time series has been treated by many, see for example [11], [12], [13]. The methods require the solution of non-linear equations and involve constrained minimization techniques. Since in the present problem quadratic functions of the moving average parameters are needed, an ad-hoc technique for the calculation of  $S_Y(f)$  given the autocorrelation function of  $\{y_t\}$  is introduced.

This method is then used to calculate some spectra of noisy signals based on theoretical values of the autocorrelation functions. This is done to demonstrate the resolution aspects of the AR and associated ARMA models for the observed signal.

Let  $\{y_t\}$  have an autocorrelation function  $\{r_t\}$ . And denote  $S_Y(z)$  resulting from equation (6) by the rational form:

$$S_Y(z) = \frac{B(z)B(z^{-1})}{D(z)D(z^{-1})} \quad (7)$$

where  $B(z) = \sum_{i=0}^{L+M} \beta_i z^i$ ,  $D(z) = \sum_{i=0}^{L+M} d_i z^i$  with  $d_0 = 1$ ,  $d_i = -\alpha_i$   $i \geq 1$  and  $\alpha_i$

the AR coefficients in the ARMA representation  $\{y_t\}$ . It is noted that, for spectral estimation, only quadratic functions of  $\beta_i$  are required. This is obvious since

$$S_Y(z) = \frac{\sum_{k=-K}^K c_k z^k}{D(z)D(z^{-1})} \quad (8)$$

where  $K = L+M$  and  $c_k$  are related to  $\beta_i$  in an obvious fashion. The scheme for computing  $S_Y(z)$  is now to first identify  $\{d_i\}$ . This can be done by a number of techniques, see for example [10] and [14]. In this paper  $\{d_i\}$  are calculated as the solution of modified Yule-Walker equations given by

$$R_K A = \rho_K \quad (9)$$

where  $A$  is given in formula (3) and

$$R_K = \begin{vmatrix} r_K & r_{K-1} & \dots & r_1 \\ r_{K+1} & r_K & \dots & r_2 \\ \vdots & & & \\ r_{2K-1} & r_{2K-2} & \dots & r_K \end{vmatrix} \text{ and } \rho_K = \begin{vmatrix} r_{K+1} \\ \vdots \\ r_{2K} \end{vmatrix}$$

Following the calculation of  $\{d_i\}$ ,  $\{c_k\}$  can be evaluated very simply by observing that by definition

$$S_Y(z) = \sum_{i=-\infty}^{\infty} r_{|i|} z^i \quad (10)$$

Equating the right-hand sides of (9) and (10)  $\{c_k\}$  can be found as

$$c_k = c_{-k} = \sum_{i=0}^K \sum_{j=0}^K \alpha_i \alpha_j r_{||i-j|-k|} \quad (11)$$

Figures [2] - [4] show the spectra of narrowband processes calculated from their theoretical autocorrelation functions using the AR(all-pole) and ARMA formulations as in formulae (2) and (9) with  $z = e^{-i2\pi f \Delta T}$ . The theoretical increase in the order of the AR spectral estimator for the same resolution of the signal in the presence of noise and interference is evident. It is noted that for strong noise and interference the ARMA model resolves the signal with much lower combined order than the AR one.

The above derivations and examples were based on exact autocorrelation functions. In practice estimates of sample correlation functions are used. Statistical properties of the rational spectral estimator introduced here are at this point unresolved. Figure [5] shows the spectra of a narrowband process observed in white Gaussian noise. The autocorrelation samples were estimated as

$$\hat{r}_i = \frac{1}{N} \sum_{t=0}^{N-1} y_t y_{t+i} \quad (12)$$

It is noted that higher order AR and ARMA models are needed for high resolution estimation of the noisy signal. An obvious explanation of this, is the fact that, for finite N, the noise-power dependent, non-zero residual sample cross-correlation between  $\{n_t\}$  and  $\{x_t\}$  and sample autocorrelation function of  $\{n_t\}$  prevent the total disentangling of the signal and noise information as for example through the modified Yule-Walker equations. This also brings out the dependence of the resolution of the above techniques on the relative signal to noise and interference power.

#### Conclusions

Spectral estimation of signals, contaminated by noise and interference, using rational spectral models was discussed. The resolution of the auto-regressive spectral estimator was related to the change in the model of the measured data. In such cases, high resolution spectral estimates require relatively high order AR(L) or an AR(K)/MA(K) model with  $K < L/2$ . An ad-hoc method was introduced for estimating the ARMA (rational) spectrum in a computationally efficient manner, based on reasonably accurate estimates of the correlation function. For relatively small sample sizes, optimum but computationally inefficient ARMA identification methods can be used off-line.

### References

1. S. A. Treter and K. Steiglitz, "Power-spectrum identification in terms of rational models", IEEE Trans. Automatic Control, vol. AC-12, pp. 185-188, April 1967.
2. T. C. Hsia and D. Landgrebe, "On a method for estimating power spectra", IEEE Trans. Instr. and Meas., Vol. IM-16, pp. 255-257, September 1967.
3. K. N. Berk, "Consistent autoregressive spectral estimates", ANN. Statist. vol. 2, pp. 489-502, May, 1974.
4. A. B. Bagherer, "Confidence intervals for regression (MEM) spectral estimates", IEEE Trans. Info. Theory, vol. IT-22, pp. 534-545, September 1976.
5. R. E. Kromer, "Asymptotic properties of the autoregressive spectral estimator", Ph.D. dissertation, Stanford Univ., Stanford, Calif., Dec. 1970.
6. R. T. Lacoss, "Data adaptive spectral analysis methods", Geophysics, vol. 36, pp. 661-675, Aug. 1971.
7. T. J. Ulrych, "Maximum entropy power spectrum of truncated sinusoids", J. Geophys. Res., vol. 77, pp. 1396-1400, Mar. 10, 1972.
8. M. Kaveh and G. R. Cooper, "An empirical investigation of the properties of the autoregressive spectral estimator", IEEE Trans. Info. Theory, vol. IT-22, pp. 313-323, May, 1976.
9. A. Nuttall, "Spectral analysis of a univariate process with bad data points, via maximum entropy and linear predictive techniques", NUSC Technical Report 5303, March 1976.
10. G. E. P. Box and G. M. Jenkins, Time Series, Forecasting and Control, San Francisco: Holden-Day, 1970.
11. J. P. Burg, "A new analysis technique for time series data", presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics.
12. R. L. Kashyap and A. R. Rao, Dynamic Stochastic Models from Empirical Data, New York: Academic Press, 1976.
13. W. Gersch, N. N. Nielsen and H. Akaike, "Maximum Likelihood estimation of structural parameters from random vibration data", Journal of Sound and Vibration, vol. 31, pp. 295-308, December, 1973.

14. W. Gersch, "Estimation of the autoregressive parameters of a mixed autoregressive moving-average time series", IEEE Trans. Automatic Control, pp. 583-588, October 1970.

Figures

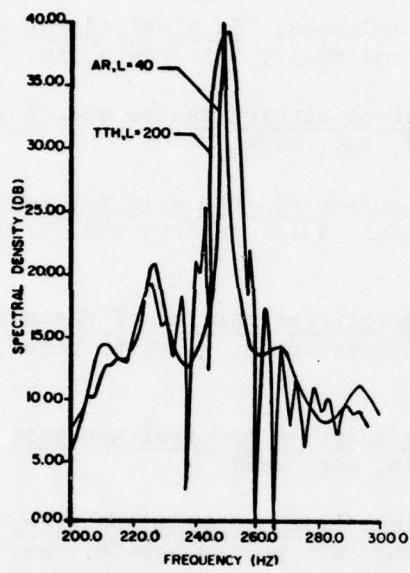


FIGURE 1. Radar power spectrum  $\Delta T = .005$  msec, 2000 samples.

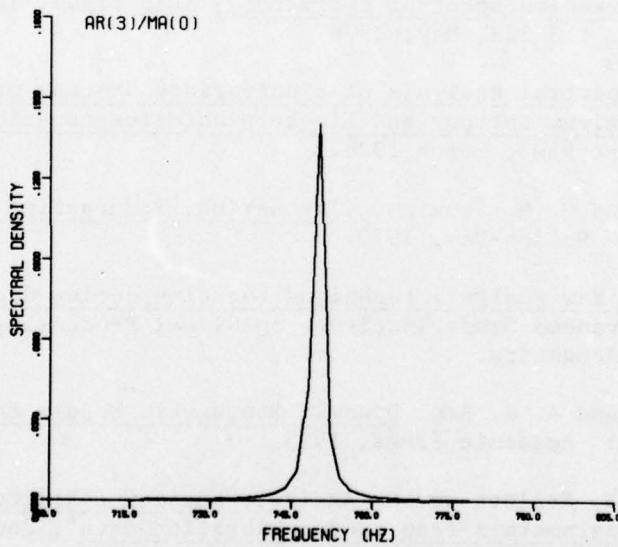
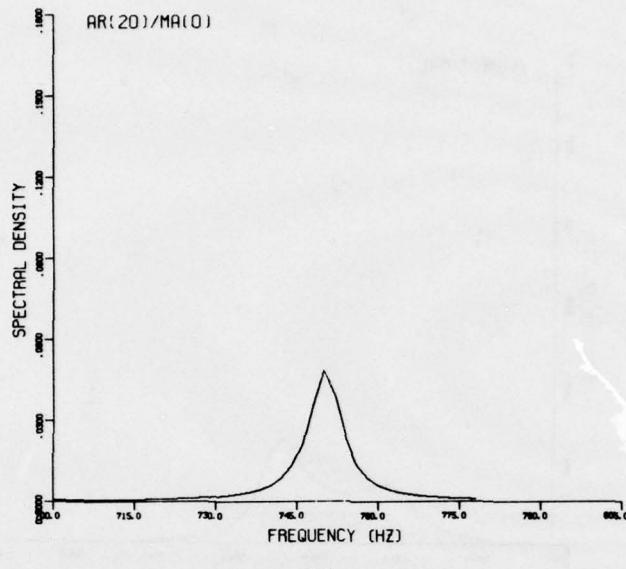
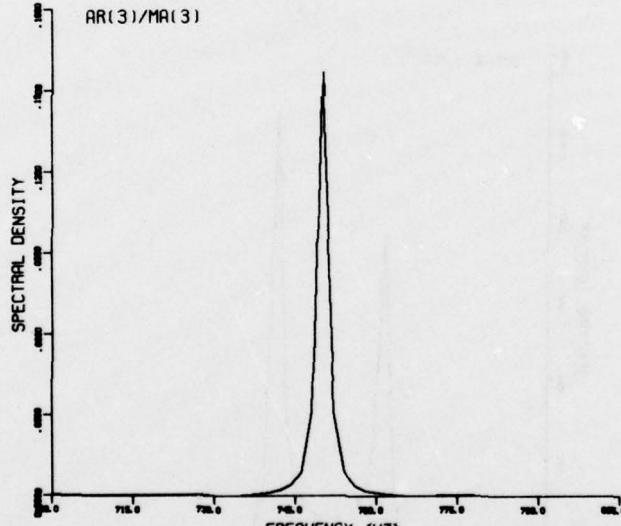


FIGURE 2. Calculated power spectrum for  $r_k = \exp(-2\pi|k|\Delta T) \cos(2\pi \times 750k\Delta T)$ ,  $\Delta T = 1/2048$ .

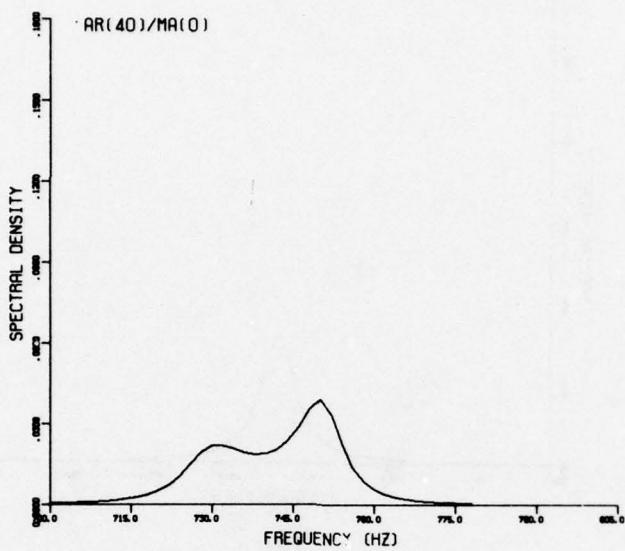


(a)

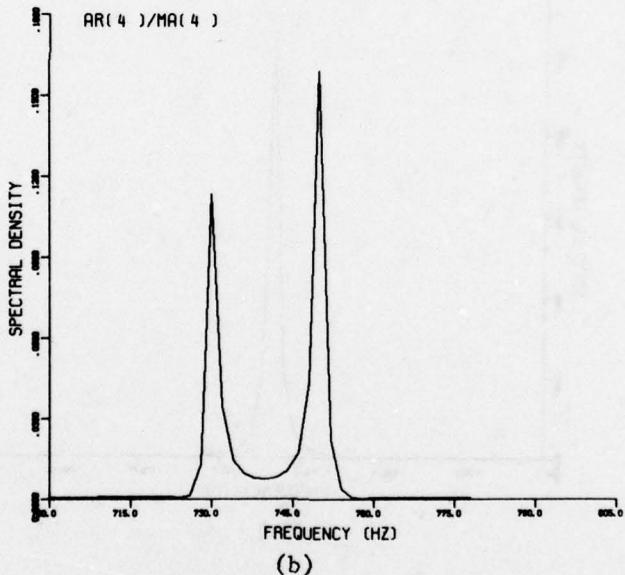


(b)

FIGURE 3. Calculated power spectrum for  $r_k = \delta_{ok} + \exp(-2\pi|k|\Delta T) \cdot \cos(2\pi x 750 k \Delta T)$ .

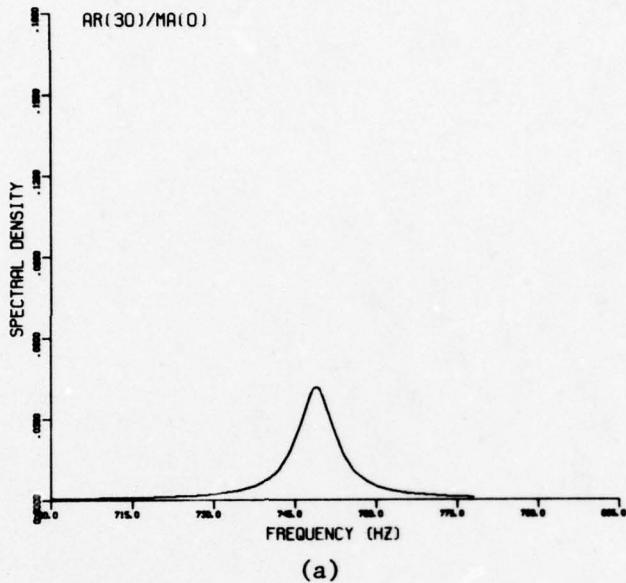


(a)

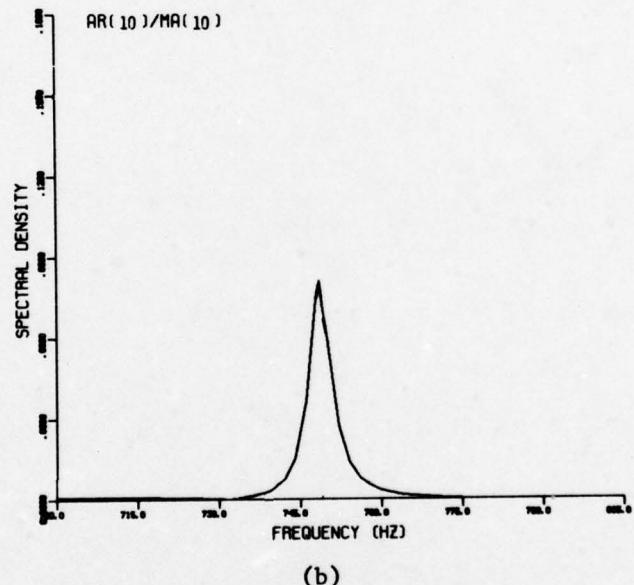


(b)

FIGURE 4. Calculated power spectrum for  $r_t = \delta + \exp(-2\pi|k|\Delta T) \cdot \cos(2\pi x 750 k \Delta T) + 0.7 \exp(-2\pi|k|\Delta T) \cdot \cos(2\pi x 730 k \Delta T)$ .



(a)



(b)

FIGURE 5. Estimated power spectrum for the signal and noise of Figure 3.  
 $N = 500$  samples.

## MAXIMUM LIKELIHOOD SPECTRAL ESTIMATION USING STATE-VARIABLE TECHNIQUES

By

Robert J. McAulay

Massachusetts Institute of Technology, Lincoln Laboratory  
Lexington, Massachusetts 02173

### ABSTRACT

The problem of spectral estimation is formulated as a fundamental problem in the detection of a random process signal in noise. The generalized likelihood ratio test for signals generated by passing white noise through a state-variable linear filter is computed and shown to depend on the energy in the prediction residuals as generated by a Kalman filter. For an all-pole filter and for large signal-to-noise ratio (SNR), it is shown that the maximum likelihood spectral estimator corresponds to solving the same normal equations as for the Maximum Entropy Method (MEM). For low SNR or for other filter characterizations the MEM technique is no longer optimum. In these cases numerical algorithms can be used to produce the maximum likelihood spectral estimates. When the signal generator consists of a linear combination of damped sinusoids the Kalman filter realization leads to an estimation algorithm that may have the ability to resolve multiple targets that are closely-spaced in frequency and contaminated by additive background noise. The essence of high-resolution is evident in the Kalman filter solution.

### INTRODUCTION

The Maximum Entropy or Linear Prediction method of spectral estimation has been widely used in those situations where high resolution is required even though the data record

lengths are short<sup>[1]</sup>. Theoretically the method is justified on the basis that for an all-pole signal model and no background noise to distort the signal the MEM technique minimizes the energy in the prediction residuals<sup>[2]</sup>. The method is then applied to problems in which the signals do not fit the all-pole model and are often contaminated by noise. In this paper an attempt is made to solve the general problem of optimally estimating the spectrum of a signal which is observed in additive noise.

### PROBLEM FORMULATION AND SOLUTION

In radar the fundamental problem is to detect the presence of a signal based on N samples of a noisy waveform. This corresponds to the hypothesis test

$$\begin{aligned} H_0: y(n) &= w(n) \\ H_1: y(n) &= s(n) + w(n) \end{aligned} \quad (1)$$

where  $w(n)$  and  $s(n)$  represent the n'th sample of the noise and signal waveforms respectively. The signal and noise are assumed to be sample functions of independent, zero mean, Gaussian random processes. The noise has variance  $\sigma_w^2$ . If an Autoregressive Moving Average<sup>w</sup>(ARMA) model is used to characterize the signal, then

$$s(n) - \sum_{k=1}^p a_k s(n-k) = \sum_{\ell=1}^q b_\ell u(n-\ell) \quad (2)$$

where the driving noise,  $u(n)$ , is zero

\*This work was sponsored by the  
Department of the Air Force.

mean, Gaussian white noise with variance  $\sigma_w^2$ . If a parallel resonator model is used to characterize the signal, then

$$s(n) = \sum_{m=1}^M s_m(n)$$

$$s_m(n) = a_{m1}s_m(n-1) + a_{m2}s_m(n-2) + a_{mo}u(n)$$

where the resonator coefficients  $(a_{m0}, a_{m1}, a_{m2})$  completely specify the amplitude, bandwidth and frequency of the  $n$ 'th resonance. Both of the above models admit to a state-variable formulation of the following form:

$$\underline{x}(n) = A \underline{x}(n-1) + b u(n) \quad (4a)$$

$$s(n) = H \underline{x}(n) \quad (4b)$$

in which  $A$  and  $b$  are directly related to the ARMA coefficients or the resonator parameters depending on whether (2) or (3) is used to characterize the signal generation process.

If these coefficients are known, then so too is the system transfer function and hence the spectral density of the signal. In general they are not known and estimates must be made of them. An optimum estimate of the power spectrum can be obtained by using the maximum likelihood method to estimate the unknown spectral coefficients. The fundamental theoretical problem is to make use of the data set  $y(1), y(2), \dots, y(N)$  to determine if a target is present and, if it is, to estimate, in an optimum way, the spectral coefficients. Schweppe [3] has shown that the likelihood functions for such a detection-estimation problem are

$$\ell_0 = N \ln (\sigma_w^2) + \frac{1}{\sigma_w^2} \sum_{n=1}^N y^2(n) \quad (5a)$$

$$\ell_1(a) = N \ln (\sigma_p^2) + \frac{1}{\sigma_p^2} \sum_{n=1}^N [y(n) - \hat{s}(n/n-1)]^2 \quad (5b)$$

where

$\ell_i = -\ln p[y(1), y(2), \dots, y(N)/H_i]$ ,  $\hat{s}(n/n-1)$  is the minimum-mean-squared-error (mmse) one-step ahead prediction of  $s(n)$  based on  $y(n-1), y(n-2), \dots$  and  $\sigma_p^2$  is the prediction error variance averaged over the ensemble of speech and noise sample functions. If the background noise variance is also assumed to be unknown then maximum likelihood estimates of  $\sigma_w^2$  and  $\sigma_p^2$  can be found by minimizing (5a) and (5b). This leads to the likelihood ratio

$$\ell(a) = \frac{\sum_{n=1}^N [y(n) - \hat{s}(n/n-1)]^2}{\sum_{n=1}^N y^2(n)} \quad (6)$$

A signal is declared present whenever  $\ell(a) < \lambda$ . Of course the spectral parameters  $a$  are unknown and must also be estimated from the data. The maximum likelihood estimates (minimum variance, unbiased for large SNR) are obtained by minimizing  $\ell(a)$ , which corresponds to minimizing the energy in the prediction residuals. Since  $\hat{s}(n/n-1)$  is the mmse one-step ahead prediction of  $s(n)$  based on  $y(n-1), y(n-2), \dots$ , it can be generated by the Kalman filter, which for the stationary case, is given by the following equations [4]

$$e(n) = y(n) - \hat{s}(n/n-1) \quad (7a) \quad e(n) = y(n) - \hat{s}(n/n-1) \quad (10a)$$

$$\hat{s}(n/n-1) = H \hat{x}(n/n-1) \quad (7b) \quad p$$

$$\hat{x}(n/n-1) = A \hat{x}(n-1) \quad (7c) \quad \hat{s}(n/n-1) = \sum_{k=1}^p a_k \hat{s}(n-k) \quad (10b)$$

$$\hat{x}(n) = \hat{x}(n/n-1) + K e(n) \quad (7d) \quad \hat{s}(n) = \hat{s}(n/n-1) + K_1 e(n) \quad (10c)$$

where the Kalman gain,  $K$ , is given by

$$K = \frac{1}{\sigma_w^2} P H^T$$

$$P = M - \frac{1}{\sigma_w^2 + H M H^T} M H^T H M$$

$$M = A P A^T + b b^T \sigma_u^2$$

#### ANALYSIS OF SPECIAL CASES

##### (a) All-Pole Model

In general computation of the Kalman gains is a difficult task. For the special case of a second order all-pole signal model, solution of equation (8) for large SNR leads to the gains

$$K_1 = \frac{\xi}{1+\xi}$$

$$K_2 = \frac{a_1 K_1 (1-K_1)}{1-a_2 (1-K_1)}$$

where  $\xi = \sigma_u^2 / \sigma_w^2$  defines the input signal-to-noise ratio. If  $\xi \rightarrow \infty$ , it follows that  $K_1 \rightarrow 1$  and  $K_2 \rightarrow 0$ . For higher order systems attempts to solve equation (8) do not lead to simple analytical expressions for the Kalman gains. It can be shown, however, that for large SNR,  $K_i = \xi / (1+\xi)$ , as in equation (9a), while  $K_i \approx 0$ ,  $i=2,3,\dots,p$ . Under these conditions the Kalman filter equations for the all-pole model can be written as follows:

where  $\hat{s}(n)$  is the MMSE estimate of  $s(n)$  based on  $y(n), y(n-1), \dots$ . The maximum likelihood estimates of the filter coefficients are obtained by minimizing

$$\ell(a_1, a_2, \dots, a_p) = \sum_{n=1}^N e^2(n) \quad (11)$$

In general it is not possible to determine the maximum likelihood estimates analytically. However, as  $\text{SNR} \rightarrow \infty$ ,  $K_1 \rightarrow 1$  and as a result  $\hat{s}(n) = y(n)$ , hence

$$\hat{s}(n/n-1) = \sum_{k=1}^p a_k y(n-k)$$

and

$$\begin{aligned} \ell(a_1, a_2, \dots, a_p) &= \sum_{n=1}^N [y(n) - \sum_{k=1}^p a_k y(n-k)]^2 \\ &\quad (12) \end{aligned}$$

which can be solved analytically. In fact the normal equations are

$$\sum_{k=1}^p a_k R_y(k-j) = R_y(j) \quad 1 \leq j \leq p \quad (13)$$

where

$$R_y(k) = \sum_{n=1}^{N-k} y(n)y(n+k) \quad (14)$$

which agree with those obtained for the maximum entropy method (MEM). As a consequence it follows that the MEM spectral estimator is optimum for the special case of an all-pole model provided the SNR is large enough.

For those situations in which non-negligible noise levels exist the MEM estimates are no longer optimum, however, the maximum likelihood estimates can be found by minimizing (11) numerically in conjunction with the dynamical constraint equations (10).

(b) Sum of Damped Sinusoids

When the received signal is modelled by a bank of resonators, solving (8) under the large SNR assumption yields the Kalman gains for the  $n^{\text{th}}$  resonator. These are

$$K_{m1} = \frac{\xi_m}{1 + \sum_{m=1}^M \xi_m} \quad (15a)$$

$$K_{m2} = \frac{a_{m1} K_{m1} (1 - K_{m1})}{1 - a_{m2} (1 - K_{m1})} \quad (15b)$$

where  $\xi_m = a_m^2 / \sigma_w^2$  represents the SNR in the  $m^{\text{th}}$  resonator channel. If the state variables for this system are taken to be  $x_{m1}(n) = s_m(n)$ ,  $x_{m2}(n) = s_m(n-1)$ , then the Kalman filter can be shown to be given by

$$e(n) = y(n) - \sum_{m=1}^M \hat{s}_m(n/n-1) \quad (16a)$$

$$\hat{s}_m(n/n-1) = \hat{x}_{m1}(n/n-1) \quad (16b)$$

$$\begin{aligned} \hat{x}_{m1}(n/n-1) &= a_{m1} \hat{x}_{m1}(n-1) \\ &+ a_{m2} \hat{x}_{m2}(n-1) \end{aligned} \quad (16c)$$

$$\hat{x}_{m2}(n/n-1) = \hat{x}_{m1}(n-1) \quad (16d)$$

$$\hat{x}_{m1}(n) = \hat{x}_{m1}(n/n-1) + K_{m1} e(n) \quad (16e)$$

$$\hat{x}_{m2}(n) = \hat{x}_{m2}(n/n-1) + K_{m2} e(n) \quad (16f)$$

The maximum likelihood estimates are obtained by minimizing the energy in the residual sequence, which must be done using numerical techniques even for the noiseless case. The Kalman filter is illustrated in Fig. 1 from which the essence of high resolution spectral estimation can be deduced. For one thing the processor uses the predictions as a means of cancelling the influence of any one sinusoid on its neighbor. Secondly, the numerical search algorithm attempts to tune a second order band-pass filter to allow each sine wave to pass in such a way that it is optimally reconstructed as  $s_m(n)$ . Estimates of the energy in each component sine wave could then be obtained by computing

$$\hat{E}_m = \sum_{n=1}^N \hat{s}_m^2(n) \quad (17)$$

As a practical matter, the resonator coefficients are related to the frequency and bandwidth of the  $m^{\text{th}}$  resonance through the equations

$$a_{m1} = 2e^{-\alpha_m} \cos \omega_m$$

$$a_{m2} = -e^{-2\alpha_m}$$

where the normalized bandwidth and resonant frequency are  $\alpha_m = \pi \beta_m / F_s$ ,  $\omega_m = 2\pi f_m / F_s$  in which  $\beta_m$  is the 3dB bandwidth in hz,  $f_m$  is the resonant frequency in hz and  $F_s$  is the sampling rate in hz. In many cases, estimates of  $\beta_m$  and  $f_m$  can be made, which provides initial estimates for the numerical search algorithm. In fact if reasonable guesses for the bandwidth can be obtained, it is then necessary to optimize only over the parameters  $a_{m1}$  which specify the frequencies of the sine waves.

In many cases, estimates of  $\beta_m$  and  $f_m$  can be made, which provides initial estimates for the numerical search algorithm. In fact if reasonable guesses for the bandwidth can be obtained, it is then necessary to optimize only over the parameters  $a_m$  which specify the frequencies of the sine waves.

#### CONCLUSIONS

The maximum likelihood method has been applied to a large class of radar detection and estimation problems in which the target signal is not known deterministically but could be modelled as the result of passing white noise through a state-variable linear filter. It was shown that optimum estimates for an all-pole signal process reduced to those obtained by the maximum entropy method provided the background noise was negligible. When the background noise level is not negligible the likelihood function can be minimized numerically so that an optimum spectral estimate can be obtained.

When the signal is characterized as the sum of damped sinusoids, which is a better representation for many situations that arise in radar practice, the maximum likelihood technique leads to an algorithm which can be solved numerically for the optimum spectral estimates. When the data can be processed off-line it is reasonable to first apply the MEM technique to get initial estimates for the frequencies and bandwidths of each of the second order Kalman filters and then fine tune the estimates using the numerical search algorithm.

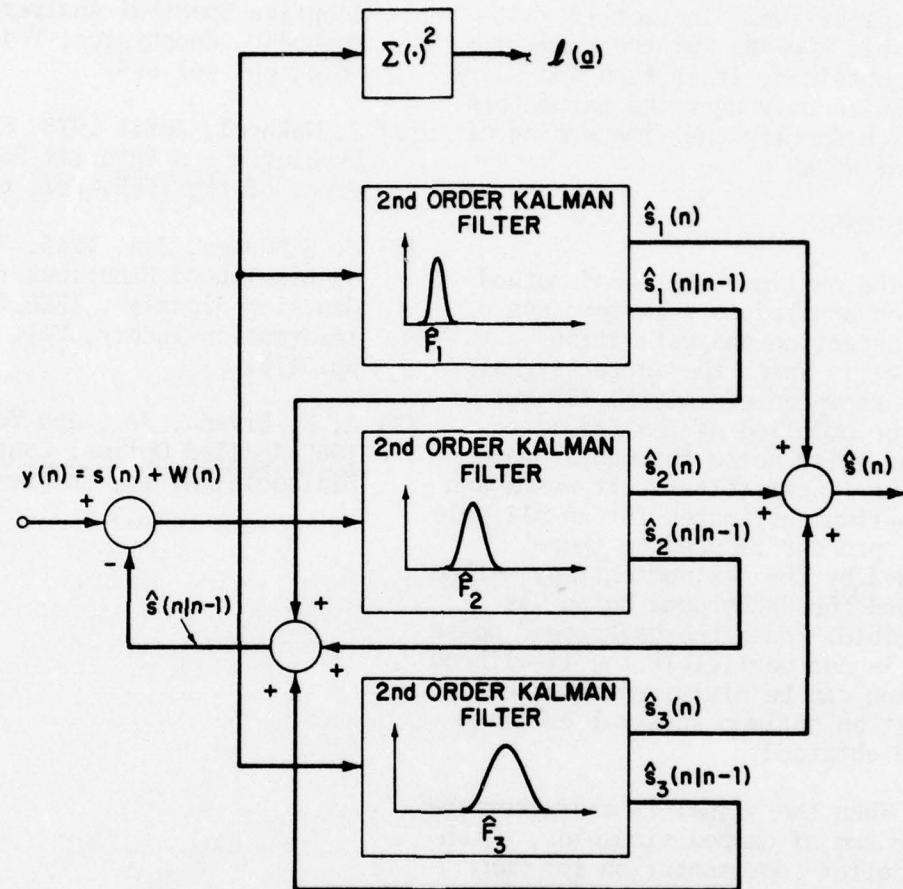
#### REFERENCES

- [1] R. T. Lacoss, Aug. 1971, "Data Adaptive Spectral Analysis Methods", Geophysics, Vol. 36, No. 4, pp. 661-615.
- [2] J. Makhoul, April 1975, "Linear Prediction: A Tutorial Review", Proc. of the IEEE, Vol. 63, No. 4.
- [3] F. Schweißpe, Jan. 1965, "Evaluation of Likelihood Functions for Gaussian Signals", IEEE Trans on Information Theory, Vol. IT-11, pp. 6170.
- [4] A. E. Bryson, Jr., and Yu-Chi Ho, 1969 Applied Optimal Control, Blaisdell, N. Y., Chapter 12.

---

"The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government."

18-00-13195 (1)



ROBUST SPEECH PROCESSOR

AN OPTIMUM FILTER FOR SPECTRAL ESTIMATION  
BASED ON A PENALTY FUNCTION APPROACH\*

RUI J. P. DE FIGUEIREDO

Department of Electrical Engineering and  
Department of Mathematical Sciences  
Rice University, Houston, Texas 77001

Abstract

A method is presented for estimating, from a finite number of observations, the autocovariance and the spectral density of a time series  $X$ . Central to our approach is the modeling of  $X$  as an autoregressive process. In this sense, our approach is akin to the maximum entropy method of Burg and the autoregressive spectral estimation method of Parzen. However, we obtain the coefficients of the filter that models the process  $X$  by a nonlinear optimization procedure which appropriately penalizes any departure of the covariance of the process which drives  $X$  from the covariance of a white noise process. Sample covariances of the process  $X$  are used in the optimization procedure. Once the filter coefficients characterizing the process  $X$  are determined, the autocovariance of the process  $X$  is estimated based on the requirement that it satisfy the Yule-Walker equations; and the spectral density of  $X$  is obtained directly from the filter coefficients and the estimated covariance values. It is believed that high resolution and accuracy in the estimation of the spectrum of  $X$  is achieved by the method.

Introduction

Since the classical contributions of Bartlett [1] [2], the developments on time series spectral estimation have been many and wide (See for example [3], [4], [5], and [6] for recent surveys). Among these developments, a point of view which appears especially promising is the one in which the series  $X$ , whose spectrum is to be estimated, is modeled as an autoregressive process. The maximum entropy method of Burg [7] and the autoregressive spectral estimation method of Parzen [3] are based on this point of view.

\*This work was supported in part by the AFOSR Grant 75-2777.

In what follows, we propose a method for the estimation of the autocovariance and spectral density of  $X$ , which is also based on the modeling of  $X$  as an autoregressive process. However, we calculate the coefficients of the filter which models  $X$  by a nonlinear optimization procedure which forces the sample covariance of the process  $Y$  that drives  $X$  (calculated from the samples of  $X$ ) to be as close to that of white noise as desired and appropriate.

Using the filter coefficients, calculated as mentioned above, the autocovariance of  $X$  is estimated by the requirement that the Yule-Walker equations be satisfied. The only assumption made in this calculation is that the covariance of  $X$  for zero lag (variance of  $X$ ) is equal to its sample value.

Finally, the spectral density of  $X$  is estimated from the filter coefficients and the estimated covariance values using a wellknown formula (equation (21)).

Our method differs from those of Burg [7] and Parzen [3] in the sense that we separate the issue of finding the optimum filter from that of estimating the covariance of the process  $X$ . Comments on the specifics and merits of the proposed approach appear in the following sections.

#### Derivation of the Optimum Filter

Let the time series  $X = \{x_1, x_2, \dots\}$  be zero-mean wide-sense-stationary, with summable and square-summable autocovariance  $R_X(\tau)$  and continuous spectral density  $f(\omega)$ .

We assume that  $X$  may be modeled as an autoregressive process of order  $p$ . We thus have

$$x_t + \sum_{j=1}^p w_j x_{t-j} = y_t, \quad t = 1, \dots, \quad (1)$$

where the filter weights  $w_j$ ,  $j = 1, \dots, p$ , are to be chosen so as to make the behavior of  $Y = \{y_t\}$  as close to that of a white noise process as possible.

Through (1), the following relationship is obtained between the autocovariances of  $Y$  and  $X$ :

$$\begin{aligned}
R_Y(\tau) &= E\{y_t y_{t+\tau}\} \\
&= E\{(x_t + \sum_{j=1}^p w_j x_{t-j})(x_{t+\tau} + \sum_{i=1}^p w_i x_{t+\tau-i})\} \\
&= E\{x_t x_{t+\tau}\} + \sum_{j=1}^p w_j E\{x_{t-j} x_{t+\tau}\} + \sum_{i=1}^p w_i E\{x_t x_{t+\tau-i}\} \\
&\quad + \sum_{j=1}^p \sum_{i=1}^p w_j w_i E\{x_{t-j} x_{t+\tau-i}\} \\
&= R_X(\tau) + \sum_{j=1}^p w_j R_X(\tau+j) + \sum_{i=1}^p w_i R_X(\tau-i) \\
&\quad + \sum_{j=1}^p \sum_{i=1}^p w_j w_i R_X(\tau+j-i). \tag{2}
\end{aligned}$$

Defining now the vectors

$$w = \text{col}(w_1, \dots, w_p), \tag{3}$$

$$r_+(\tau) = \text{col}(R_X(\tau+1), \dots, R_X(\tau+p)), \tag{4}$$

$$r_-(\tau) = \text{col}(R_X(\tau-1), \dots, R_X(\tau-p)), \tag{5}$$

and the  $p \times p$  matrix  $\tilde{R}_X(\tau)$  whose  $j$ <sup>th</sup> element is  $R_X(\tau+j-i)$ , we may express (2) in the concise form

$$R_Y(\tau) = R_X(\tau) + w^T (r_+(\tau) + r_-(\tau)) + w^T \tilde{R}_X(\tau) w, \tag{6}$$

where the superscript T denotes the transpose.

To make Y approximate white noise behavior, we appropriately minimize  $R_Y(\tau)$  for  $\tau \neq 0$  over all w. Specifically, we select the minimization criterion\*

$$\begin{aligned}
J(w) &= \sum_{\tau=1}^p g(\tau) R_Y^2(\tau) = \sum_{\tau=1}^p \left\{ g(\tau) [R_X(\tau) + w^T (r_+(\tau) + r_-(\tau)) \right. \\
&\quad \left. + w^T \tilde{R}_X(\tau) w]^2 \right\}. \tag{7}
\end{aligned}$$

(From now on, we will abbreviate  $\sum_{\tau=1}^p$  simply by  $\sum$ ). Above,  $g(\tau)$  is a nonnegative even penalty function selected to express the extent to which it is appropriate for Y to satisfy the white noise hypothesis. There are two sources

\*Note that for  $g(\tau) = 1$ , (7) reduces to a least squares criterion.

of error which contribute to the broadening of the autocorrelation of  $Y$  at the origin. One type of error occurs if we use too low a value for the order  $p$  of our model (1) of the process  $X$ . The other source of error arises from the fact that we use sample covariances of  $X$  rather than the true covariances, in the minimization of (7). So if the number  $N$  of samples of  $X$  is small, values of  $R_Y(\tau)$  slightly away from the origin should not be penalized too heavily.

The expression that we suggest for  $g(\tau)$  is

$$g(\tau) = (\epsilon\tau)^{2n}, \quad (8)$$

where  $\epsilon$  is a positive number and  $n$  a nonnegative integer.  $g(\tau)$  is equal to unity for  $\tau = (1/\epsilon)$ , and for large  $n$ ,  $g(\tau)$  is nearly zero for  $\tau < (1/\epsilon)$  and rises sharply to high values as  $\tau$  increases from  $(1/\epsilon)$ . So by appropriately selecting the parameters  $\epsilon$  and  $n$ , one is able to tune the minimization criterion (7) to the model and to the data.

Now, the gradient of (7) is

$$\begin{aligned} \nabla_w J(w) &= \sum_{\tau} \left\{ g(\tau) [R_X(\tau) + w^T(r_+(\tau) + r_-(\tau)) + w^T \tilde{R}_X(\tau) w] \right. \\ &\quad \cdot [r_+(\tau) + r_-(\tau) + 2 \tilde{R}_X(\tau) w] \} \\ &= \sum_{\tau} \left\{ g(\tau) [R_X(\tau) (r_+(\tau) + r_-(\tau)) + ((r_+(\tau) + r_-(\tau)) \right. \\ &\quad \cdot (r_+(\tau) + r_-(\tau))^T + 2 R_X(\tau) \tilde{R}_X(\tau) w \\ &\quad + (2w^T(r_+(\tau) + r_-(\tau)) \tilde{R}_X(\tau) w + w^T \tilde{R}_X(\tau) w \\ &\quad \cdot (r_+(\tau) + r_-(\tau))) + 2 w^T \tilde{R}_X(\tau) w \tilde{R}_X(\tau) w] \} \quad (9) \end{aligned}$$

The above expression can be conveniently written as

$$\nabla_w J(w) = h(w), \quad (10)$$

$h(w)$  being a  $p$ -vector with components defined by

$$h_i(w) = a^i + \sum_j b_{j,i}^i w_j + \sum_j \sum_k c_{jk,i}^i w_j w_k + \sum_j \sum_k \sum_l d_{jkl,i}^i w_j w_k w_l, \quad i = 1, \dots, p, \quad (11)$$

where the values of the constants  $a^i$ ,  $b_j^i$ ,  $c_{jk}^i$ ,  $d_{jkl}^i$  are obtained by carrying out the summations with respect to  $\tau$  in (9).

The following propositions are clear:

Proposition 1: If  $w^*$  is a minimizer of  $J$  then

$$h(w^*) = 0. \quad (12)$$

Proposition 2: If in addition to (12),  $w^*$  satisfies

$$\frac{\partial h_i}{\partial w_j}(w^*) > 0, \quad i, j = 1, \dots, p, \quad (13)$$

then  $w^*$  is a local minimizer of  $J$ .

Algorithm: The above provides the basis for obtaining the optimum filter weight vector  $w^*$  by implementation of a standard minimization algorithm (for example, based on the gradient method or the Newton-Raphson method) on a digital computer. For this purpose, we use sample covariances for the various covariance terms appearing in (7). Since the value of  $p$  is in general not too high, the algorithm should be easy to run.

#### Estimation of the Autocovariance and the Spectrum

With the optimum filter weight vector calculated as in the preceding section, we proceed to estimate the autocovariance  $R_X(\tau)$  as follows.

We equate  $R_X(0)$  to its sample value, that is\*

$$R_X(0) = \frac{1}{N} \sum_{t=1}^N x_t^2. \quad (14)$$

The remaining values of  $R_X(\tau)$  are obtained by requiring that the Yule-Walker equations be satisfied. Using our previous notation, these equations take the form

$$w^{*T} r_-(\tau) = -R_X(\tau), \quad \tau = 1, \dots, p. \quad (15)$$

Above, since  $R_X(-\tau) = R_X(\tau)$  and the values of the optimum

\*For the sake of generality, we allow the number  $N$  of samples of  $X$  to be different from  $p$ , if necessary.

filter weight vector  $w^*$  and  $R_X(0)$  (according to (14)) are known, we have exactly  $p$  equations in the  $p$  unknowns  $R_X(1), \dots, R_X(p)$ , which constitute the components of the vector  $r_+(0)$ . So they can be solved for  $r_+(0)$ .

For example, in the case in which  $p = 4$ , we may express (15) in terms of a single matrix equation in the form

$$W r_+(0) = q, \quad (16)$$

where

$$W = \begin{pmatrix} (1+w_2^*) & w_3^* & w_4^* & 0 \\ (w_1^*+w_3^*) & (1+w_4^*) & 0 & 0 \\ (w_2^*+w_4^*) & w_1^* & 1 & 0 \\ w_3^* & w_2^* & w_1^* & 1 \end{pmatrix}, \quad (17)$$

and

$$q = -R_X(0) w^*. \quad (18)$$

Hence,  $r_+(0)$  is obtained by inverting (16):

$$r_+(0) = W^{-1} q. \quad (19)$$

To get the spectrum, first calculate  $R_Y(0)$  from (6):

$$R_Y(0) = R_X(0) + 2w^{*T} r_+(0) + w^{*T} \tilde{R}_X(0) w^*, \quad (20)$$

where we have used the fact that  $r_-(0) = r_+(0)$ . Note also that all the quantities in the right side of (20) have been calculated.

The spectral density is finally obtained in terms of the above quantities by the formula

$$f(\omega) = \frac{R_Y(0)}{2\pi \left| 1 + \sum_{k=1}^p w_k^* e^{-i\omega k} \right|^2}. \quad (21)$$

### Conclusion

A method has been described for the estimation of the auto covariance and the spectral density of a stationary time series. The proposed technique is a variant of the methods of Burg [7] and Parzen[3]. The proposed framework allows one to incorporate realistic constraints in the estimation procedure.

### References

1. Bartlett, M.S., 1948, "Smoothing Periodograms from Time Series with Continuous Spectra", Nature, 161, pp.666-668.
2. Bartlett, M.S., 1950, "Periodogram Analyses and Continuous Spectra", Biometrika, 37, pp. 1-16.
3. Parzen, E., 1972, "Some Recent Advances in Time Series Analysis", in Statistical Models and Turbulence (M. Rosenblatt and C. Van Atta, eds.), Springer-Verlag, Berlin and New York.
4. Kanasewich, E.R., 1973, "Time Sequence Analysis in Geophysics", The University of Alberta Press, Edmonton, Alberta, Canada.
5. Koopmans, L.H., 1974, "The Spectral Analysis of Time Series", Academic Press, New York.
6. Box, G.E.P. and Jenkins, G.M., 1976, "Time Series Analysis Forecasting and Control" (Revised Edition), Holden-Day, San Francisco.
7. Burg, J.P., 1967, "Maximum Entropy Spectral Analysis", Paper presented at the 37th Annual Int. SEG Meeting, Oklahoma, October 1967; Preprint - Texas Instruments, Dallas.

A SOLUTION TO THE PROBLEM OF SPONTANEOUS LINE SPLITTING  
IN MAXIMUM ENTROPY POWER SPECTRUM ANALYSIS OF COMPLEX SIGNALS

PAUL F. FOUGERE

Air Force Geophysics Laboratory  
Hanscom AFB, Bedford, MA 01731

Abstract

Under certain conditions, Burg's maximum entropy spectra of real signals in the presence of additive noise show either spontaneous line splitting (low noise levels) or appreciable frequency shifting (at moderate noise levels). This difficulty arises because an unnecessary constraint is imposed during the minimization of the mean error power. A similar problem arises when the Burg technique is applied to complex signals. This paper presents a solution to this latter problem.

Introduction

In a recent paper, Fougere et al [1], have shown that under certain conditions the maximum entropy method using Burg's prediction error coefficients produces power spectra which display spurious line splitting in the presence of very low noise. In a second paper hereinafter called Paper II, Fougere [2] showed that this type of splitting occurred only if the noise level were sufficiently low; when the noise level is gradually increased the spectrum is broadened and the multiple peaks coalesce into a single peak shifted substantially away from the correct value. In Paper II, Fougere presented a solution to that problem and showed that using that solution, splitting was cured in the low noise case and shifting was reduced considerably in the moderate noise cases. These two papers treated only the case of real input time series, and therefore real prediction error coefficients.

For general information on the maximum entropy technique see the PhD thesis by Burg [3] and papers by Smylie et al [4], Ulrych and Bishop [5], Ulrych and Clayton [6], and Radoski et al [7].

The present paper extends these results to the case of a complex time series requiring complex prediction error coefficients. The first observations of spontaneous splitting with

a complex signal in the presence of very low noise were by R.W. Herring (private communication). This paper will follow closely the structure given in the "Detailed Mathematics" section of paper II. Each of the equations in paper II will be written in the appropriate complex form. The equation numbers will be the same.

#### Detailed Mathematics

We are given an  $n$ -point sample  $(x_1, x_2, \dots, x_n)$  of complex numbers  $x_i$ , measured at equally spaced values of a single real independent variable, usually considered to be time. Define an  $(m+1)$  point prediction error filter (PEF)  $(1, g_{m1}, g_{m2}, \dots, g_{mm})$  where each  $g_{ij}$  is a complex variable, such that the  $k$ 'th prediction errors are:

$$\epsilon_{1k} = \sum_{i=0}^m x_{k+m-i} g_{mi} \quad (1)$$

$$\epsilon_{2k} = \sum_{i=0}^m x_{k+i} g_{mi}^*$$

$$k = 1, 2, 3, \dots, i-m$$

where  $g_{mi}^*$  is the complex conjugate of  $g_{mi}$ ,  $g_{m0} \equiv 1$ , and  $\epsilon_{1k}$  and  $\epsilon_{2k}$  are the forward and backward prediction errors, respectively.

Now the mean square prediction error, or mean error power, in both time directions is:

$$P_m = 0.5 (n-m)^{-1} \sum_{s=1}^2 \sum_{k=1}^{n-m} \epsilon_{sk} \epsilon_{sk}^* \quad (2)$$

If the PEF's (with leading "1" suppressed) of all orders  $1, 2, \dots, m$  are gathered in one complex matrix  $G_m$ , we may write:

$$G_m = \begin{pmatrix} g_{11} & & \\ g_{21} & g_{22} & \\ \vdots & & \\ \vdots & & \\ g_{m1} & g_{m2} & \dots & g_{mm} \end{pmatrix} \quad (3)$$

The generalization of the Levinson Algorithm is given by

$$g_{jk} = g_{j-1,k} + g_{jj}^* g_{j-1,j-k} \quad (4)$$

This simple, two term formula allows the off diagonal elements of the  $j$ th row of  $G_m$  to be determined wherever the diagonal elements  $\{g_{jj}, j=1,m\}$  are known.

Burg has shown that if these diagonal elements (also called reflection coefficients) all lie in the range  $|g_{jj}| < 1$ , then the PEF is minimum phase, that is its Z transform has all its zeroes outside the unit circle.

In order to enforce this condition we set

$$g_{jj} = U \sin \theta_j e^{i\phi_j} \quad (5)$$

where  $\theta_j$  and  $\phi_j$  are any real angles and  $U$  is a positive constant slightly less than unity. The discussion in paper II on the significance of  $U$ , can be carried over unchanged to the present case of complex input data. Briefly,  $U$  is adjusted so that all of the roots of the Z transform of the PEF all lie outside the unit circle and none lie on it.

We now follow a method used by Cain et al [8]; variations in  $\theta_j$  and  $\phi_j$  are written

$$\begin{aligned} \theta_j &= \theta_j^0 + \Delta\theta_j \\ \phi_j &= \phi_j^0 + \Delta\phi_j \end{aligned} \quad (6)$$

Now expand the prediction errors  $\epsilon_{sk}$  in a Taylor series about  $\theta_j^0$

and  $\phi_j^o$  and retain only the first two orders:

$$\varepsilon_{sk} = \varepsilon_{sk}^o + \sum_{j=1}^m \left( \frac{\partial \varepsilon_{sk}}{\partial \theta_j} \Delta \theta_j + \frac{\partial \varepsilon_{sk}}{\partial \phi_j} \Delta \phi_j \right) \quad (7)$$

Substitute (7) into (2) to get:

$$P_m = 0.5(n-m)^{-1} \sum_{s=1}^2 \sum_{k=1}^{n-m} \left| \varepsilon_{sk}^o + \sum_{j=1}^m \left( \frac{\partial \varepsilon_{sk}}{\partial \theta_j} \Delta \theta_j + \frac{\partial \varepsilon_{sk}}{\partial \phi_j} \Delta \phi_j \right) \right|^2 \quad (8)$$

Set  $\partial P_m / \partial \Delta \theta_\alpha = 0$  to find the minimum error power and then rearrange the resulting equations to get:

$$\begin{aligned} & \sum_{j=1}^m \sum_{s=1}^2 \sum_{k=1}^{n-m} \left[ \left( \frac{\partial \varepsilon_{sk}}{\partial \theta_j} \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} + \frac{\partial \varepsilon_{sk}^*}{\partial \theta_j} \frac{\partial \varepsilon_{sk}}{\partial \theta_\alpha} \right) \Delta \theta_j \right. \\ & \quad \left. + \left( \frac{\partial \varepsilon_{sk}}{\partial \phi_j} \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} + \frac{\partial \varepsilon_{sk}^*}{\partial \phi_j} \frac{\partial \varepsilon_{sk}}{\partial \theta_\alpha} \right) \Delta \phi_j \right] \\ & = - \sum_{s=1}^2 \sum_{k=1}^{n-m} \left( \varepsilon_{sk}^o \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} + \varepsilon_{sk}^{o*} \frac{\partial \varepsilon_{sk}}{\partial \theta_\alpha} \right) \end{aligned}$$

There are three expression in parentheses. Note that each has the form  $A + A^* = 2 \operatorname{Re}\{A\}$ . Thus the equation becomes

$$\begin{aligned} & \sum_{j=1}^m \sum_{s=1}^2 \sum_{k=1}^{n-m} \left[ \operatorname{Re} \left( \frac{\partial \varepsilon_{sk}}{\partial \theta_j} \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} \right) \Delta \theta_j \right. \\ & \quad \left. + \operatorname{Re} \left( \frac{\partial \varepsilon_{sk}}{\partial \phi_j} \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} \right) \Delta \phi_j \right] = - \sum_{s=1}^2 \sum_{k=1}^{n-m} \operatorname{Re} \left( \varepsilon_{sk}^o \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} \right) \quad (9a) \end{aligned}$$

Similarly when we set  $\partial P_m / \partial \Delta \phi_\alpha = 0$  we get

$$\sum_{j=1}^m \sum_{s=1}^2 \sum_{k=1}^{n-m} \left[ \operatorname{Re} \left( \frac{\partial \varepsilon_{sk}}{\partial \theta_j} \frac{\partial \varepsilon_{sk}^*}{\partial \phi_\alpha} \right) \Delta \theta_j + \operatorname{Re} \left( \frac{\partial \varepsilon_{sk}}{\partial \phi_j} \frac{\partial \varepsilon_{sk}^*}{\partial \phi_\alpha} \right) \Delta \phi_j \right] = - \sum_{s=1}^2 \sum_{k=1}^{n-m} \operatorname{Re} \left( \varepsilon_{sk}^0 \frac{\partial \varepsilon_{sk}^*}{\partial \phi_\alpha} \right) \quad (9b)$$

where, in both (9a) and (9b):  $\alpha = 1, 2, \dots, m$ .

The  $2m$  equations (9) are now linear in the corrections  $\Delta \theta_j$  and  $\Delta \phi_j$  and can therefore be solved by standard matrix methods. The corrections are then substituted into (6) and the process is repeated until the corrections  $\Delta \theta_j$  and  $\Delta \phi_j$  become sufficiently small.

In order to find the derivatives:  $\partial \varepsilon_{sk}/\partial \theta_j$ ,  $\partial \varepsilon_{sk}^*/\partial \theta_j$ ,  $\partial \varepsilon_{sk}/\partial \phi_j$  and  $\partial \varepsilon_{sk}^*/\partial \phi_j$ , we differentiate (1) as follows:

$$\begin{aligned} \frac{\partial \varepsilon_{1k}}{\partial \theta_j} &= \sum_{i=1}^m x_{k+m-i} \frac{\partial g_{mi}}{\partial \theta_j} ; \quad \frac{\partial \varepsilon_{1k}}{\partial \phi_j} = \sum_{i=1}^m x_{k+m-i} \frac{\partial g_{mi}}{\partial \phi_j} \\ \frac{\partial \varepsilon_{2k}}{\partial \theta_j} &= \sum_{i=1}^m x_{k+i} \frac{\partial g_{mi}^*}{\partial \theta_j} ; \quad \frac{\partial \varepsilon_{2k}}{\partial \phi_j} = \sum_{i=1}^m x_{k+i} \frac{\partial g_{mi}^*}{\partial \phi_j} \end{aligned} \quad (10)$$

and four equations resulting from (10) by complex conjugation, for example

$$\frac{\partial \varepsilon_{1k}^*}{\partial \theta_j} = \sum_{i=1}^m x_{k+m-i}^* \frac{\partial g_{mi}^*}{\partial \theta_j}$$

Next we rewrite eq. (5) and its complex conjugate

$$\begin{aligned} g &= U \sin \theta e^{i\phi} \\ g^* &= U \sin \theta e^{-i\phi} \end{aligned} \quad (A)$$

where we have temporarily dropped the subscripts for convenience. Equations (A) may be inverted to yield:

$$\begin{aligned}\theta &= \theta(g, g^*) \\ \phi &= \phi(g, g^*)\end{aligned}\tag{B}$$

If we treat  $g$  and  $g^*$  as independent variables, we can express the partial derivatives with respect to  $\theta$  and  $\phi$  as follows:

$$\begin{aligned}\frac{\partial}{\partial \theta} &= \frac{\partial g}{\partial \theta} \frac{\partial}{\partial g} + \frac{\partial g^*}{\partial \theta} \frac{\partial}{\partial g^*} \\ \frac{\partial}{\partial \phi} &= \frac{\partial g}{\partial \phi} \frac{\partial}{\partial g} + \frac{\partial g^*}{\partial \phi} \frac{\partial}{\partial g^*}\end{aligned}\tag{C}$$

Substituting  $e^{i\phi} = \cos \phi + i \sin \phi$  in (A) and performing the differentiation indicated in (C) we arrive at:

$$\begin{aligned}\frac{\partial}{\partial \theta} &= U \cos \theta [\cos \phi (\frac{\partial}{\partial g} + \frac{\partial}{\partial g^*}) + i \sin \phi (\frac{\partial}{\partial g} - \frac{\partial}{\partial g^*})] \\ \frac{\partial}{\partial \phi} &= U \sin \theta [-\sin \phi (\frac{\partial}{\partial g} + \frac{\partial}{\partial g^*}) + i \cos \phi (\frac{\partial}{\partial g} - \frac{\partial}{\partial g^*})]\end{aligned}\tag{D}$$

Write

$$D_j^\pm \equiv \frac{\partial}{\partial g_{jj}} \pm \frac{\partial}{\partial g_{jj}^*} \tag{E}$$

Then

$$\begin{aligned}\frac{\partial}{\partial \theta_j} &= U \cos \theta_j (\cos \phi_j D_j^+ + i \sin \phi_j D_j^-) \\ \frac{\partial}{\partial \phi_j} &= U \sin \theta_j (-\sin \phi_j D_j^+ + i \cos \phi_j D_j^-)\end{aligned}\tag{i1}$$

Note that because  $g_{jj}$  and  $g_{ii}^*$  are independent variables,  $\partial g_{jj}/\partial g_{ii}^* = 0$  and  $\partial g_{jj}^*/\partial g_{ii} = 0$ , for any values of  $i$  and  $j$ .

We now apply (E) to the Levinson Algorithm (4). The result is

$$D_j^\pm g_{ik} = D_j^\pm g_{i-1,k} + g_{i-1,i-k}^* \delta_{ij} + g_{ii} D_j^\pm g_{i-1,i-k}^* \tag{12}$$

Where the middle term arises because

$$D_j^{\pm} g_{ii} = \left( \frac{\partial}{\partial g_{jj}} \pm \frac{\partial}{\partial g_{jj}^*} \right) g_{ii} = \delta_{ij} \quad (F)$$

If we take the complex conjugate of equation (E) we get:

$$(D_j^{\pm})^* = \frac{\partial}{\partial g_{jj}^*} \pm \frac{\partial}{\partial g_{jj}} = \pm D_j^{\pm} \quad (G)$$

Therefore, anytime we need derivatives of any terms in  $g^*$  we can write

$$D_j^{\pm} g_{km}^* = \pm (D_j^{\pm} g_{km})^* \quad (H)$$

Finally, the derivatives with respect to  $\theta_j$  and  $\phi_j$  are determined by substituting (12) into (11).

The gradient of  $P_m$  with respect to the independent variables  $\theta_K$  and  $\phi_K$  is written:

$$\nabla_{\theta, \phi} P_m = \sum_{\alpha=1}^m \left( \frac{\partial P_m}{\partial \theta_\alpha} \hat{\theta}_\alpha + \frac{\partial P_m}{\partial \phi_\alpha} \hat{\phi}_\alpha \right) \quad (13)$$

where  $\hat{\theta}_\alpha$  and  $\hat{\phi}_\alpha$  are unit vectors.

Applying (13) to (2) yields:

$$\begin{aligned} \frac{\partial P_m}{\partial \theta_\alpha} &= 0.5(n-m)^{-1} \sum_{s=1}^2 \sum_{k=1}^{n-m} \varepsilon_{sk} \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} + \varepsilon_{sk}^* \frac{\partial \varepsilon_{sk}}{\partial \theta_\alpha} \\ &= (n-m)^{-1} \sum_{s=1}^2 \sum_{k=1}^{n-m} \operatorname{Re} \left( \varepsilon_{sk} \frac{\partial \varepsilon_{sk}^*}{\partial \theta_\alpha} \right) \end{aligned}$$

Thus,

$$\frac{\partial P_m}{\partial \theta_\alpha} = -(n-m)^{-1} \times (\text{right side of (9a)}) \quad (14a)$$

Similarly,

$$\frac{\partial P_m}{\partial \theta_\alpha} = -(n-m)^{-1} \times (\text{right side of (9b)}) \quad (14b)$$

This completes the formal derivation. Clearly these results must be programmed for a computer before we can test the method. When such a program has been written and checked out, it will be made available on request to seriously interested scientists.

#### References

1. Fougere, P.F., E.J. Zawalick, and H.R. Radoski, 1976, "Spontaneous line splitting in maximum entropy power spectrum analysis", Phys. Earth Planet. Interiors, 12, P. 201-207.
2. Fougere, P.F., 1977, "A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis", J. Geophys. Res., 82, P. 1051-1054.
3. Burg, J.P., 1975, "Maximum entropy spectral analysis", Ph.D. thesis, Stanford Univ., Stanford, CA.
4. Smylie, D.E., G.K.C. Clarke and T.J. Ulrych, 1973, "Analysis of irregularities in the earth's rotation", Methods Comput. Phys., 13, P. 391-430.
5. Ulrych, T.J. and T.N. Bishop, 1975, "Maximum entropy spectral analysis and autoregressive decomposition", Rev. Geophys. Space Phys., 13, P. 183-200.
6. Ulrych, T.J. and R.W. Clayton, 1976, "Time series modelling and maximum entropy", Phys. Earth Planet. Interiors, 12, P. 188-200.
7. Radoski, H.R., P.F. Fougere and E.J. Zawalick, 1975, A comparison of power spectral estimates and application of the maximum entropy method", J. Geophys. Res., 80, P. 619-625.
8. Cain, J.C., S.J. Hendricks, R.A. Langel, and W.V. Hudson, 1967, "A proposed model for the international geomagnetic reference field - 1965", J. Geomagn. Geoelec., 19, P. 335-355.

## ADAPTIVE EXTRAPOLATION AND HIDDEN PERIODICITIES

A. PAPOULIS and C. CHAMZAS

Department of Electrical Engineering  
Polytechnic Institute of New York  
Route 110, Farmingdale, N. Y. 11735

### 1. Introduction

An important problem in many applications is the determination of the frequency components of a signal

$$f(t) = \sum_{i=1}^m c_i e^{j\omega_i t} \quad (1)$$

in terms of the segment .

$$w_1(t) = \begin{cases} f(t)+n(t) & |t| < T \\ 0 & |t| > T \end{cases} \quad (2)$$

of  $f(t)$  containing the noise component  $n(t)$ . The signal  $f(t)$  is not known for every  $t$  for a variety of reasons:

The signal  $f(t)$  can be written as a sum of exponentials for a limited time only (Voice; non-stationary processes).

The available time of observation is limited (sun spots; weather trends).

Measurements are limited by instrument constraints (Michelson interferometer; band-limited channels).

The unknown frequencies  $\omega_i$  and coefficients  $c_i$  can be determined simply with ordinary Fourier transforms if the time of observation  $2T$  is large compared to all the periods  $T_i = 2\pi/\omega_i$ . This is not, however, the case if  $T$  is of the order  $T_i$ , particularly if the noise component of  $w_1(t)$  is not negligible. In this paper, we present a method which, as we hope to

show, is reliable even in such extreme cases.

The method involves only FFT and it is based on earlier results dealing with the problem of extrapolating band-limited functions [1, 2]. We review for easy reference the relevant parts of these results.

## 2. Extrapolation of band-limited functions

Consider a function  $f(t)$  with Fourier transform  $F(\omega)$  such that

$$F(\omega) = 0 \quad |\omega| > \sigma \quad (3)$$

We form the function

$$w_1(t) = \begin{cases} f(t) & |t| < T \\ 0 & |t| > T \end{cases} \quad (4)$$

obtained by truncating  $f(t)$  as in Fig.1. We shall determine  $f(t)$  in terms of  $w_1(t)$  by numerical iteration.

First step. We compute the Fourier transform  $W_1(\omega)$  of  $w_1(t)$ , form the function

$$F_1(\omega) = \begin{cases} W_1(\omega) & |\omega| < \sigma \\ 0 & |\omega| > \sigma \end{cases} \quad (5)$$

Compute the inverse transform  $f_1(t)$  of  $F_1(\omega)$ , and form the function

$$w_2(t) = \begin{cases} w_1(t) = f(t) & |t| < T \\ f_1(t) & |t| > T \end{cases} \quad (6)$$

and its Fourier transform  $W_2(\omega)$ .

This completes the first step of the iteration (Fig.1).

nth step. We form the function

$$F_n(\omega) = \begin{cases} W_n(\omega) & |\omega| < \sigma \\ 0 & |\omega| > \sigma \end{cases} \quad (7)$$

where  $W_n(\omega)$  is the function obtained at the end of the preceding step. We compute the inverse transform  $f_n(t)$  of  $F_n(\omega)$ , form the function

$$w_{n+1}(t) = \begin{cases} f(t) & |t| < T \\ f_n(t) & |t| > T \end{cases} \quad (8)$$

and compute its Fourier transform  $W_{n+1}(\omega)$ .

If  $f(t)$  is approximated by  $f_n(t)$ , the resulting mean-square error is given by

$$E_n = \int_{-\infty}^{\infty} [f(t) - f_n(t)]^2 dt = \frac{1}{2\pi} \int_{-\sigma}^{\sigma} |F(\omega) - F_n(\omega)|^2 d\omega \quad (9)$$

We maintain that this error decreases twice at each iteration step. Indeed,

$$E_n = \int_{|t| < T} [f(t) - f_n(t)]^2 dt + \int_{|t| > T} [f(t) - f_n(t)]^2 dt$$

But [see (8) and (7)]

$$\begin{aligned} \int_{|t| > T} [f(t) - f_n(t)]^2 dt &= \int_{-\infty}^{\infty} [f(t) - w_{n+1}(t)]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega) - W_{n+1}(\omega)|^2 d\omega \\ &= \frac{1}{2\pi} \int_{|\omega| > \sigma} |F(\omega) - W_{n+1}(\omega)|^2 d\omega + \frac{1}{2\pi} \int_{-\sigma}^{\sigma} |F(\omega) - F_{n+1}(\omega)|^2 d\omega \end{aligned}$$

Hence,

$$\begin{aligned} E_n - E_{n+1} &= \int_{|t|<T} \left[ f(t) - f_n(t) \right]^2 dt + \frac{1}{2\pi} \int_{|\omega|>\sigma} |F(\omega) - W_{n+1}(\omega)|^2 d\omega \\ &= \int_{|t|<T} \left[ f(t) - f_n(t) \right]^2 dt + \frac{1}{2\pi} \int_{|\omega|>\sigma} |F(\omega) - W_{n+1}(\omega)|^2 d\omega \end{aligned} \quad (10)$$

In [1] and [2] we show that  $f_n(t) \rightarrow f(t)$  as  $n \rightarrow \infty$ . This is not true if the given segment  $w_1(t)$  of  $f(t)$  is noisy<sup>n</sup> as in (2). In this case, a satisfactory estimate of  $f(t)$  can be found by early termination of the iteration. [2]

Note. From (10) it follows that the mean-square error  $E_n$  is a monoton decreasing function and since it is positive, it tends to a limit. This does not prove the convergence of (9) because the limit need not be zero. It shows, however, that

$$E_n - E_{n+1} \rightarrow 0 \quad n \rightarrow \infty$$

Hence,

$$\int_{|t|<T} \left[ f(t) - f_n(t) \right]^2 dt \rightarrow 0 \quad n \rightarrow \infty \quad (11)$$

Although the functions  $f(t)$  and  $f_n(t)$  are band-limited, (11) does not imply that  $f(t) \rightarrow f_n(t)$  because there is no lower bound on the energy concentration of band-limited functions in a finite interval [1, 3]. For example, the prolated spheroidal functions  $\varphi_n(t)$  are band-limited, their energy equals one but their energy concentration in the interval  $(-T, T)$  tends to zero as  $n \rightarrow \infty$ . This is the case because the eigenvalues  $\lambda_n$  of the underlying integral equation tend to zero as  $n \rightarrow \infty$ .

We mention without elaboration that, in the discrete version of the problem, the convergence of the iteration can be deduced from (11) under suitable conditions. The reason is that the corresponding eigenvalues are finitely many, therefore, they have a positive minimum.

### 3. Adaptive extrapolation

The preceding method was based on the assumption that the unknown function  $f(t)$  is band-limited. This information was used to reduce the error in the estimation of  $f(t)$  twice at each iteration step. The speed of iteration can be increased and the effects of noise can be reduced if additional a priori information about  $f(t)$  is available. Suppose, for example, that the size of the band of  $F(\omega)$  is known but its precise location is unknown. We then choose a constant  $\sigma$  sufficiently large for  $F(\omega)$  to vanish outside the integral  $(-\sigma, \sigma)$  and proceed as in Sec. 2. As the iteration progresses, the form of  $W_n(\omega)$  suggests appropriate reduction of the assumed band of  $f(t)$ .

The adaptive extrapolation method is particularly effective if  $f(t)$  is a sum of exponentials as in (1). In this case,  $F(\omega)$  consists of impulses (lines) as in Fig. 2:

$$F(\omega) = 2\pi \sum_{i=1}^m c_i \delta(\omega - \omega_i) \quad (12)$$

and our problem is to determine their locations  $\omega_i$  and amplitudes  $c_i$  in terms of the known segment  $w_n(t)$  of  $f(t)$ .

To solve this problem, we select a constant  $\sigma$  larger than the largest possible value of  $\omega_i$  and we proceed with the iteration until  $W_n(\omega)$  takes significant values only in a subset  $B_n$  of the band  $(-\sigma, \sigma)$  of  $f(t)$  (Fig. 3). This suggests that the unknown frequencies are in  $B_n$ . When this is observed, the function  $F_n(\omega)$  of the nth iteration step is obtained from the following modification of (7):

$$F_n(\omega) = \begin{cases} W_n(\omega) & \omega \in B_n \\ 0 & \omega \in \bar{B}_n \end{cases} \quad (13)$$

(Fig. 3) In the above,  $B_n$  is the set of points such that  $W_n(\omega)$  exceeds a threshold level  $\epsilon_n$

$$W_n(\omega) = \begin{cases} > \epsilon_n & \omega \in B_n \\ < \epsilon_n & \omega \in \bar{B}_n \end{cases} \quad (14)$$

and  $\bar{B}_n$  its complement. The process is repeated until  $W_n(\omega)$  approach the unknown spectrum. This can be checked by comparing the inverse  $w_n(t)$  of  $W_n(\omega)$  with the known segment of  $f(t)$ .

Notes. As the following examples show, the unknown frequencies can be found even if the data are noisy and the constant  $T$  is smaller than the smallest period  $T_i$ .

The choice of the threshold level  $\epsilon_n$  is dictated by two conflicting factors: For a speedy convergence and reduction of the noise component,  $\epsilon_n$  must be large. It must be sufficiently small so that all frequency components of  $F(\omega)$  are in the set  $B_n$ . Thus  $\epsilon_n$  is small at first and it increases as the iteration progresses.

The accuracy of the method depends on the number  $m$  of the unknown components and their relative locations and amplitudes. If some components are small compared to the maximum  $c_i$ , it is possible that they could be lost. However, if the noise is sufficiently small, they can be recovered by subtracting the significant components and repeating the iteration.

A priori knowledge of the number  $m$  of the unknown frequencies is useful but not essential.

If, at the  $n$ th iteration step, all frequency components of  $f(t)$  are in the set  $B_n$ , then the resulting mean-square error reduction is given by (10) mutatis mutandis.

#### 4. Illustrations

We conclude with a digital implementation of the above method. The computations were performed with a PDP 11 minicomputer (single precision) and the FFT size was  $N = 256$ . The known segment  $w_1(t)$  of  $f(t)$  contains the first 30 points. In figure 4 the unknown signal consists of 3 cosine waves. Determination of their amplitudes and frequencies is not apparent from the Fourier transform  $W_1(\omega)$ , figure 4b, of the given segment. The resolution is improved considerably within a few steps of the iteration. In the 12th step the three cosine terms have been revealed, figure 4c, and in the 63rd step complete estimation of frequencies and amplitudes have been achieved. (figure 4d).

In the next example, figure 5, the same signal is used but we added 10% white noise uniformly distributed. The signal is again recovered completely in 100 steps. In figure 6 the unknown signal consisting of two exponentials has been corrupted by 20% white noise. To recover the signal the iteration needs 30 steps.

Thus, as we see from these illustrations the unknown frequencies can be found even if the data are noisy and the constant  $T$  is smaller than the smallest period in  $f(t)$ .

#### References

- [1] A. Papoulis, Signal Analysis, McGraw-Hill, New York
- [2] , "A New Algorithm in Spectral Analysis and Band-limited Extrapolation", IEEE Trans. Circuits Syst., vol. CAS-22, no. 9, pp. 735-742, September 1975.
- [3] D. Slepian, H.J. Landau, and H.O. Pollack, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty Principle I and II", Bell Syst. Tech. J., vol. 40, no. 1, 1961.
- [4] A. Papoulis and M. Bertran, Digital Filtering and Prolate Functions", IEEE Trans. Circuit Theory, vol. CT-19, pp. 674-681, November 1972.

Figures

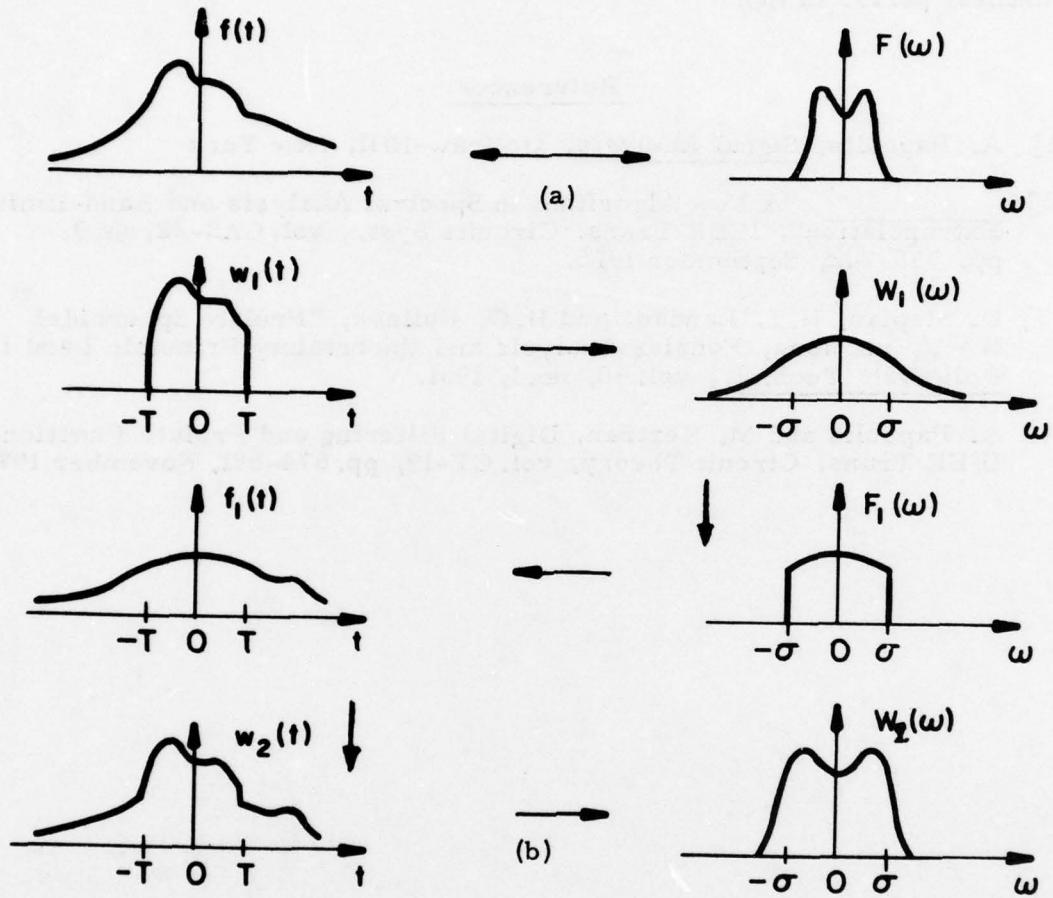


FIGURE 1. (a) The unknown signal  $f(t)$  and its Fourier transform  $F(\omega)$   
 (b) First iteration step starting with known segment  $w_1(t)$ .

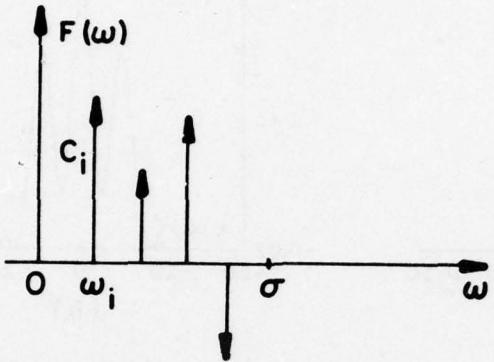


FIGURE 2. Fourier transform of the unknown signal.

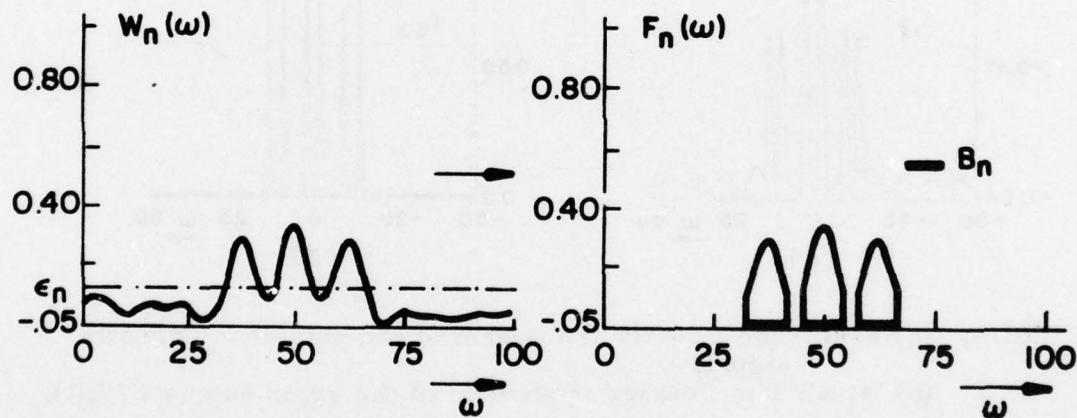


FIGURE 3. Truncation of  $W_n(\omega)$  below a threshold level  $\epsilon_n$  yielding  $F_n(\omega)$ .

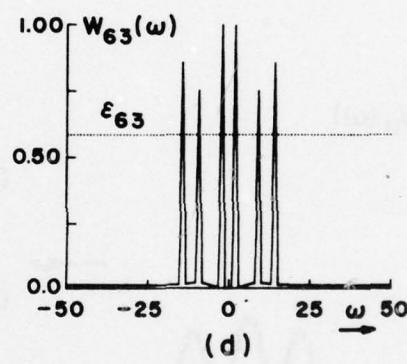
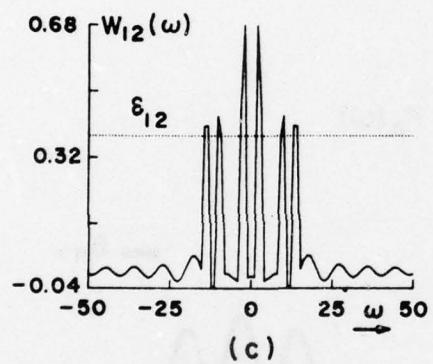
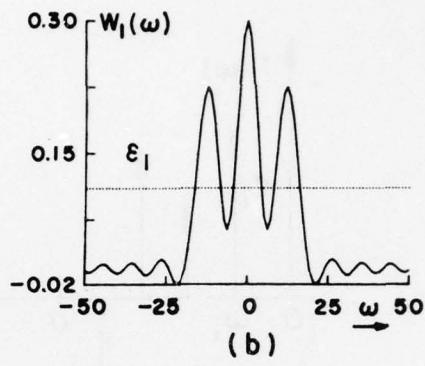
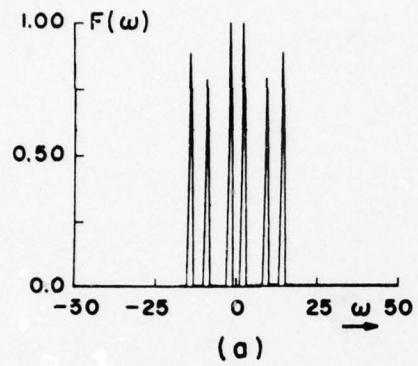


FIGURE 4. (a)  $F(\omega)$ : The Fourier transform of the noiseless unknown signal.  
 (b)  $W_1(\omega)$ : The Fourier transform of the given segment  $W_1(t)$ .  
 (c), (d): The result of the 12th and 63rd iteration  
 $(\epsilon_n$  indicates the threshold level at the nth step)

AD-A054 650      ROME AIR DEVELOPMENT CENTER GRIFFISS AFB N Y  
PROCEEDINGS OF THE RAD C SPECTRUM ESTIMATION WORKSHOP HELD ON 24--ETC(U)  
MAY 78

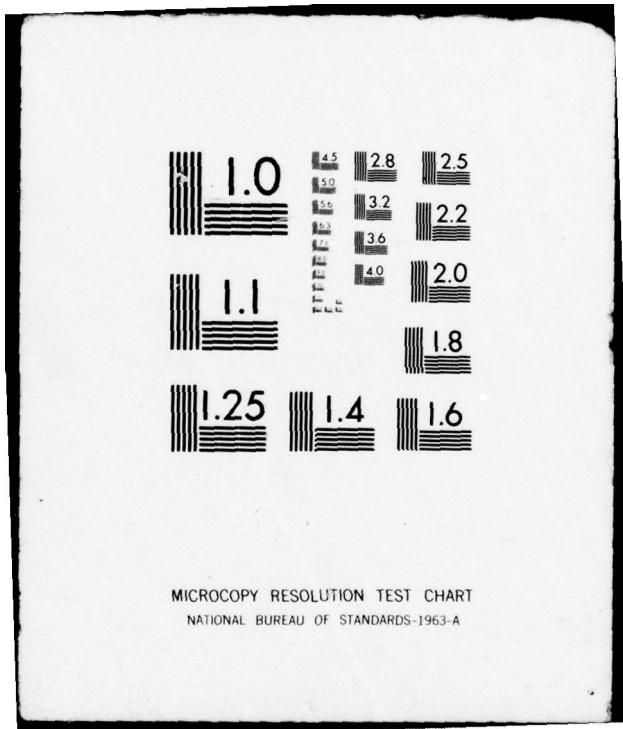
F/G 14/2

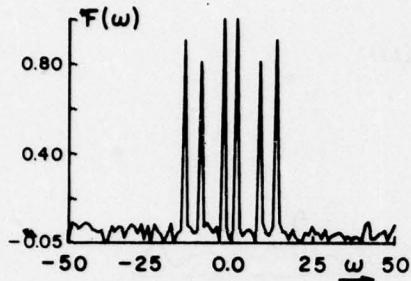
UNCLASSIFIED

NL

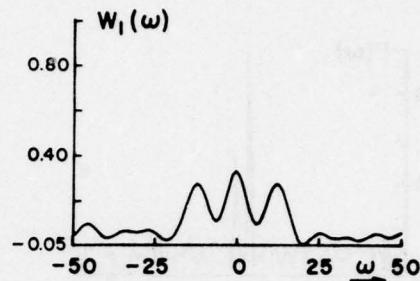
2 OF 4  
AD-A054650



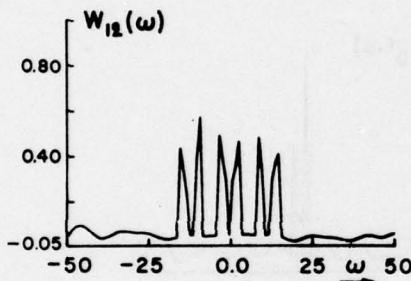




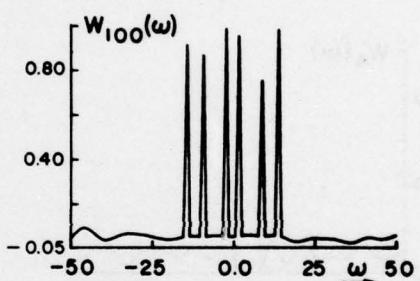
(a)



(b)

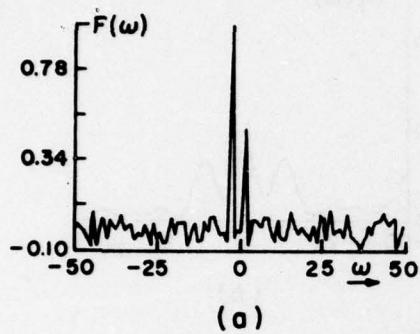


(c)

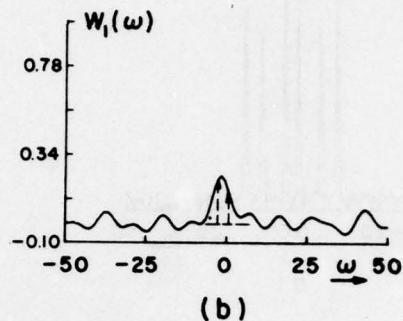


(d)

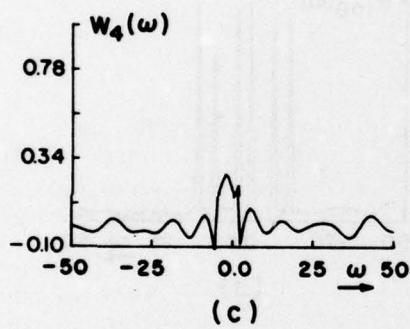
FIGURE 5. (a)  $F(\omega)$ : The Fourier transform of the unknown signal  $f(t)$ .  
 (b)  $W_1(\omega)$ : The Fourier transform of the given segment  $w_1(t)$ .  
 (c), (d): The result of the 12th and 100th iteration.



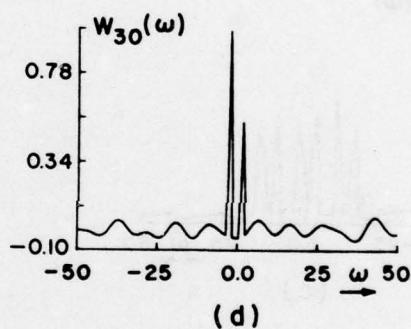
(a)



(b)



(c)



(d)

FIGURE 6. (a)  $F(\omega)$ : Unknown signal consisting of two impulses contaminated by noise.

(b)  $W_1(\omega)$ : Fourier transform of the known segment  $w_1(t)$ .

(c),(d): The result of the 4th and 30th iteration.

## **Recursive Spectral Estimation of $\alpha$ - Stationary Processes\***

**M. Morf, D. T. L. Lee**

**Information Systems Laboratory  
Dept. of Electrical Engineering  
Stanford University, Stanford, CA. 94305**

### **Abstract**

We have recently developed the mathematical tools to treat a certain class of non-stationary processes that are of great interest. Most current spectral estimation techniques make implicitly the somewhat tenuous assumption that the observed processes are stationary; however, the most frequent case of sine waves in (stationary) white noise is really nonstationary. Current approaches typically randomize the phase in order to achieve at least a stationary ensemble. Although the non-randomized processes are nonstationary, they possess close to stationarity features. One of them is the fact that their covariances exhibit a finite so-called "displacement rank"  $\alpha$ . Matrices with low displacement rank also appear in the problem of fitting autoregressive (AR) models to sample covariances. The significance of this low displacement rank is that it takes  $O(n^2\alpha)$  instead of  $O(n^3)$  operations to solve normal equations associated with fitting  $n$ -th order AR models to such covariances.

---

\* This work was supported by the Air Force Office of Scientific Research, Air Force Systems Command under Contract AF44-620-74-C-0068; the Joint Services Electronics Program under Contract N00014-75-C-0601; National Science Foundation under Contract NSF Eng75-18952; and by ARPA through the use of the Stanford Artificial Intelligence Laboratory facilities.

## I. Introduction

Spectral estimation has in the past received a lot of attention; however in recent years, new approaches to spectral analysis have been developed and applied with considerable success in many areas, such as speech processing, geophysical data analysis, as well as sonar and radar signal enhancement. The bases for such developments have been available for some time, for instance, in statistics through the work of Parzen [Par] and in geophysics the work of Burg [Bur1]. Burg's motivation was to choose the spectrum that corresponds to the most random or the most unpredictable time series whose covariance function agrees with the known values, say  $\{R_0, R_1, \dots, R_N\}$ , thus leading to the name maximum entropy method (MEM). This procedure has been shown (see e.g. Van Den Bos [VDB]) to be equivalent to computing the spectrum via the least-squares fitting of an autoregressive (AR) (or all-pole) model to the given covariance, which was the procedure suggested by Parzen. An AR process is one in which the best least-squares prediction is given by a weighted sum of previous values.

The problem of fitting an AR model to a given covariance function leads to the solution of a set of equations known as the Yule-Walker equations. These equations can be solved in an efficient recursive way by an algorithm first presented by Levinson (1942) and Durbin (1960) for the scalar case and then extended to the vector case by Whittle (1963), Wiggins and Robinson (1965), and Burg (1967) (see the references in [WR], [K-S74], [Mo1]). We refer to it as the LWR algorithm. It recursively fits AR schemes of increasing orders. At each step, the forward and backward one-step linear predictors (say of order  $p$ ) are found as a linear combination of the predictors of order  $p-1$ .

In the scalar case, it is known (see e.g. [Bur2], [MVK]) that the LWR algorithm produces a sequence of reflection coefficients, having magnitude less than or equal to one, which have a one-to-one correspondence with the given covariance. Therefore, these numbers can be used to parametrize the spectrum of a process directly. In addition, the reflection coefficients determine, through the LWR algorithm, the parameters of the fitted AR scheme, or in other words the time

domain model. Recently we have shown in [MVLK] that for the vector case, we can also obtain a sequence of matrix reflection coefficients, which have a one-to-one correspondence with the given matrix covariance functions. This is achieved by using a suitably normalized forms of the LWR algorithm, a simple case is given in Section II.

In practice, we do not have available the exact values of the covariance function but instead we are only given a finite segment of a time series  $\{y_t\}$ . A classical method of fitting an AR scheme to the data is then to estimate the lagged covariances by, say,

$$R_k = \frac{1}{N-k} \sum_{t=1}^{N-k} y_t y_{t+k}$$

and then to use them in the Yule-Walker equation. Burg's technique is to directly estimate the reflection coefficients from the data, ignoring the covariance function. These estimates are obtained by minimizing the sum of the squares of the forward and backward one-step prediction errors (or innovations) [Bur1], [Bur2].

In most of the classical methods as well as Parzen's (and Burg's) technique, an implicit assumption is made that the observed data are samples of stationary processes. For instance in Parzen's (and Burg's) method, it is assumed that the AR model that is to be fitted to the data represents a stationary process. However, in most practical cases, the observed date are not necessarily AR and often also not very stationary. More generally, one would like to fit a model that reflects as closely as possible the underlying physics of the problem. For example, if (white) measurement noise is added to the output of an AR model driven by some (white) input process, a model of the observations that is driven by a single (white) noise, the so-called "innovations model", will have a moving average part and is therefore a more general autoregressive moving average (ARMA) model. Another example is the case where a sine wave with fixed initial phase is represented as a undriven second order AR process with fixed initial conditions. The output of this process is clearly non-stationary. In linear problems the "natural" functions are complex exponentials, leading to rational spectra and thus to ARMA type models. The

common case of "spectral lines" in white noise unfortunately is strictly speaking non-stationary and not AR. Thus in general we need also to include spectral zeros, or from a process point of view include the so-called MA (moving average) models. The presence of zeros leads unfortunately to difficulties, as in many other areas such as communications and control. It is possible however to embed the ARMA problem in a (multichannel) AR model as shown in [Mo2]. We will concentrate here on the simpler AR case.

In recent work [FMKL], [FKM], [Ka] and [MK], we have shown that the above cases fall into a class of processes which we called  $\alpha$ -stationary.. These processes can be characterized by an index of the "distance from stationarity" of the process. It turns out that matrices with low displacement rank also appear in the problem of fitting autoregressive (AR) models to sample covariances. The significance of this low displacement rank is that it takes  $O(n^2\alpha)$  instead of  $O(n^3)$  operations to solve normal equations associated with fitting n-th order AR models to such covariances.

In Section II we will illustrate the basic ideas of the (normalized) LWR algorithm, and in Section III we discuss some extensions to non-stationary processes as well as methods to treat the more general ARMA models.

## II. The Normalized LWR Algorithm

Suppose we are given the  $m \times m$  matrices

$$R_n = E \{ y_t y_{t+n}^T \} , \quad |n| \leq N$$

where  $\{y_t\}$  is an  $m$ -vector stationary random process, so that  $R_{-n} = R_n^T$ . Then the so-called maximum entropy extension of the sequence  $\{R_n, |n| \leq N\}$  is defined by the expressions

$$\begin{aligned}
 R(z) &= \sum_{n=-\infty}^{\infty} R_n z^{-n} \\
 &= A_N^{-1}(z) R_N^\epsilon A_N^{-T}(z^{-1}) \\
 &= B_N^{-1}(z) R_N^r B_N^{-T}(z^{-1})
 \end{aligned} \tag{1}$$

where  $A_N(z)$  and  $B_N(z)$  are transfer functions of the so-called forward and backward prediction filters, and  $R_N^\epsilon$ ,  $R_N^r$  are the respective prediction-error (or innovation) variances. These quantities are defined by the equations (see e.g., [WR], [K-S74]).

$$\begin{bmatrix} I & A_{N,1} & \dots & A_{N,N} \\ & \ddots & & \\ B_{N,N} & \dots & B_{N,1} & I \end{bmatrix} R_N = \begin{bmatrix} R_N^\epsilon & 0 & \dots & 0 \\ 0 & \dots & 0 & R_N^r \end{bmatrix} \tag{2}$$

where

$$R_N = \begin{bmatrix} R_0 & R_{-1} & \dots & R_{-N} \\ R_1 & \dots & \dots & \dots \\ R_N & \dots & \dots & R_0 \end{bmatrix}$$

and

$$A_N(z) = I + A_{N,1}z^{-1} + \dots + A_{N,N}z^{-N} \tag{3}$$

$$B_N(z) = B_{N,N} + B_{N,N-1}z^{-1} + \dots + I z^{-N} \tag{4}$$

Note that (2) is just the (Yule-Walker) equation obtained in minimizing  $E\{\epsilon_{N,t}^\top \epsilon_{N,t}\}$  and  $E\{r_{N,t}^\top r_{N,t}\}$  where

$$\epsilon_{N,t} = y_t + A_{N,1}y_{t-1} + \dots + A_{N,N}y_{t-N} \tag{5}$$

$$r_{N,t} = B_{N,N}y_t + \dots + B_{N,1}y_{t-N+1} + y_{t-N} \tag{6}$$

are respectively the forward and backward prediction errors. These equations can

be solved in an efficient recursive manner by using the LWR algorithm ([WR], [K-S74]). In [MVK], [MVLK] we have shown that the LWR can be put in the normalized form given below. We only give the main results here leaving detailed descriptions to [MVK], [MVLK].

The normalized LWR algorithm can be expressed in the following compact form

$$\begin{bmatrix} \tilde{A}_{n+1}(z) \\ \tilde{B}_{n+1}(z) \end{bmatrix} = \Theta_{n+1}(z) \begin{bmatrix} \tilde{A}_n(z) \\ \tilde{B}_n(z) \end{bmatrix}, \quad (7)$$

where

$$\Theta_{n+1}(z) = \begin{bmatrix} P_{n+1} & 0 \\ 0 & Q_{n+1} \end{bmatrix}^{-1/2} \begin{bmatrix} I & -\rho_{n+1} \\ -\rho_{n+1}^T & I \end{bmatrix} \begin{bmatrix} zI & 0 \\ 0 & I \end{bmatrix}, \quad (8)$$

and also

$$A_n(z) = (\tilde{A}_{n,0})^{-1} \tilde{A}_n(z) \quad (9)$$

$$B_n(z) = (\tilde{B}_{n,0})^{-1} \tilde{B}_n(z) \quad (10)$$

$$(R_n^\epsilon)^{1/2} = P_0^{1/2} \dots P_n^{1/2} \quad (11)$$

$$R_0^r = R_0^\epsilon = P_0 = Q_0 = R_0 \quad (12)$$

$$\tilde{A}_{0,0} = \tilde{B}_{0,0} = (R_0^r)^{-1/2} \quad (13)$$

Here the  $\rho_{n+1}$  are the reflection coefficients given by

$$\rho_{n+1} = (R_n^\epsilon)^{-1/2} \Delta_{n+1} (R_n^r)^{-T/2} \quad (14)$$

$$\Delta_{n+1} = R_{n+1} + A_{n,1} R_n + \dots + A_{n,n} R_1. \quad (15)$$

Furthermore, if we define

$$\Theta^{(n+1)}(z) = \Theta_{n+1}(z) \Theta_n(z) \Theta_{n-1}(z) \dots \Theta_0, \quad (16)$$

where

$$\Theta_0 = \begin{bmatrix} P_0 & 0 \\ 0 & Q_0 \end{bmatrix}^{-1/2}$$

then we also have

$$\begin{bmatrix} \tilde{A}_{n+1}(z) \\ \tilde{B}_{n+1}(z) \end{bmatrix} = \Theta^{(n+1)}(z) \begin{bmatrix} I \\ I \end{bmatrix}. \quad (17)$$

The estimated multichannel spectrum is then given by

$$R(z) = \tilde{A}_N^{-1}(z) \tilde{A}_N^{-T}(z^{-1}) = \tilde{B}_N^{-1}(z) \tilde{B}_N^{-T}(z^{-1}). \quad (18)$$

### III. Extensions to Nonstationary and ARMA Models.

In [MLVN], [MVL] we have developed recursive algorithms for finding linear predictors, i.e. fitting AR type models, to data that are not stationary. These algorithms are recursive in time and order, they are also well suited to track time-varying parameters of the underlying AR model. The reflection coefficients mentioned earlier turned out to be a natural parametrization of these algorithms, leading to realizations in so-called ladder canonical forms. These forms also have several other nice features such as lowest computational and storage requirements as well as a "stability by inspection" property. Detailed descriptions of these ladder forms are given in [MLVN], [MVL] and [Mo2]. In the Appendix we present some examples out of a set of extensive simulations we performed, in order to demonstrate their tracking behavior of time-varying parameters.

The extension of AR type modeling methods to the ARMA case has been demonstrated in the system identification literature, a discussion of several such methods can be found in [GP], [SLG], [MLK]. A direct way of demonstrating that such an extension is possible can be found in [MLVN], [Mo2], where it is shown that considering the joint process of observation driving noise the ARMA modeling problem can be embedded in an AR modeling problem given the joint process.

### References

- [Bur1] Burg, J. P., "Maximum Entropy Spectral Analysis," presented at the 37th Annual Meeting of Society Explor. Geophys., Oklahoma City, Oklahoma, 1967.
- [Bur2] Burg, J. P., "Maximum Entropy Analysis," Ph.D. Dissertation, Stanford University, Stanford, Ca., 1975.
- [FKM] Friendlander, B., T. Kailath and M. Morf, "A Modified Displacement Rank and Some Applications," Proc. 1977 IEEE Conf. on Decision and Control, pp.958-961, New Orleans, Louisiana, Dec. 7-9, 1977.
- [FMKL] Friendlander, B., M. Morf, T. Kailath and L Ljung, "Levinson- and Chandrasekhar-Type Equations for a General Discrete-Time Linear Estimation Problem," Proc. 1976 IEEE Conf. on Decision and Control, Florida, Dec. 1976.
- [GD] Goodwin, G. and R. Payne, *Dynamic System Identification*, Academic Press, New York, 1977.
- [Ka] Kailath, T., "Some New Methods for Solving Linear Equations," Trans. 22nd Conf. Army Mathematicians, pp.211-225, ARO Rep. 77-1, 1977.
- [K-S74] Kailath, T., "A View of Three Decades of Linear Filtering Theory," IEEE Trans. on Information Theory, Vol. IT-20, No.2, pp. 145-181, March 1974.
- [MK] Morf, M. and T. Kailath, "Recent Results in Least-Squares Estimation Theory," Ann. Economic Social Meas., Vol.6, No.3, pp.261-274, 1977.
- [MLK] Morf, M., L. Ljung and T. Kailath, "Fast Algorithms for Recursive Identification," Proc. 1976 IEEE Conf. on Decision and Control, Florida, Dec. 1976.
- [MLNV] Morf, M., D. T. Lee, J. Nickolls and A. Vieira, "A Classification of Algorithms for ARMA Models and Ladder Realizations," Proc. 1977 IEEE Acoustics, Speech, and Signal Processing Conf., Hartford, Conn., May 9-11, 1977.

- [Mo1] Morf, M., "Fast Algorithms for Multivariable Systems," Ph.D. Dissertation, Stanford University, Stanford, Ca. 1975.
- [Mo2] Morf, M., "Ladder Forms in Estimation and System Identification," Proc. 11th Asilomar Conf. on Circuits and Systems, Dec. 1977.
- [MVK] Morf, M., A. Vieira, and T. Kailath, "Covariance Characterization by Partial Autocorrelation Matrices," submitted for publication in 1976, to appear in *Annals of Statistics*, May 1978.
- [MVL] Morf, M., A. Vieira, and D. T. Lee, "Ladder Forms for Identification and Speech Processing," Proc. Conf. on Decision and Control, New Orleans, Louisiana, Dec. 7-9, 1977.
- [MVLK] Morf, M., A. Vieira, D. T. Lee and T. Kailath, "Recursive Maximum Entropy Spectral Estimation," IEEE Trans. Geoscience Electronics, Vol. 16, Apr. 1978.
- [Par] Parzen, E., "Multiple Time Series Modeling," in *Multivariate Analysis 2*, ed. P. R. Krishnaian, Academic , N.Y., 1969, pp.389-409.
- [SLG] Soderstrom, T., L.Ljung, and I. Gustavsson, "A Comparative Study of Recursive Identification," Rep. 7427, Div. of Auto. Contr., Lund Inst. of Tech., 1974.
- [VDB] Van Den Bos, A., "Alternate Interpretation of Maximum Entropy Spectral Analysis," IEEE Trans. Inform. Theory, Vol. IT-17, pp.493-494, 1971.
- [WR] Wiggins R, A, and E. A. Robinson, "Recursive Solution to the Multichannel Filtering Problem," J. Geophys. Res., Vol.70, 1965, pp.1885-1891.

#### Appendix

The data used in this simulation was generated by an 8th order AR process with time-varying (piecewise constant) reflection coefficients, driven by the sum of a Gaussian white noise and a non-Gaussian impulse train. The pulses occur at the step changes in the parameters. Figure 1 shows the data generated by such a model. Figures 2 and 3 show the actual and estimated parameters of the model. The simulations show the tracking behaviour that the ladder form is capable of.

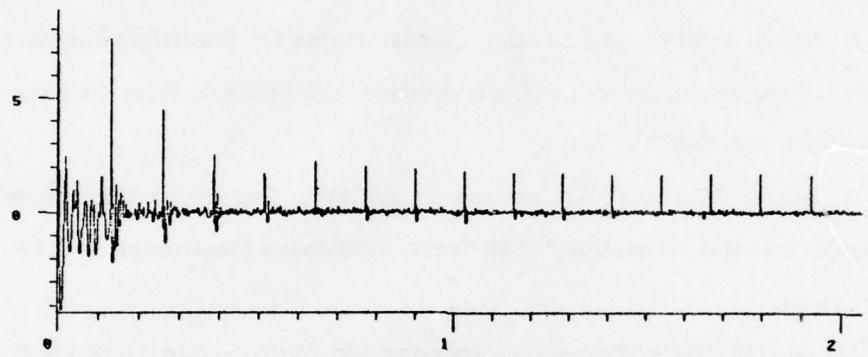


Figure 1. An 8th order AR process with time-varying  
(piecewise constant) reflection coefficients  
that converge exponentially to zero.  
Process was driven by sum of Gaussian white  
noise and an impulse train. There are 2000  
samples.

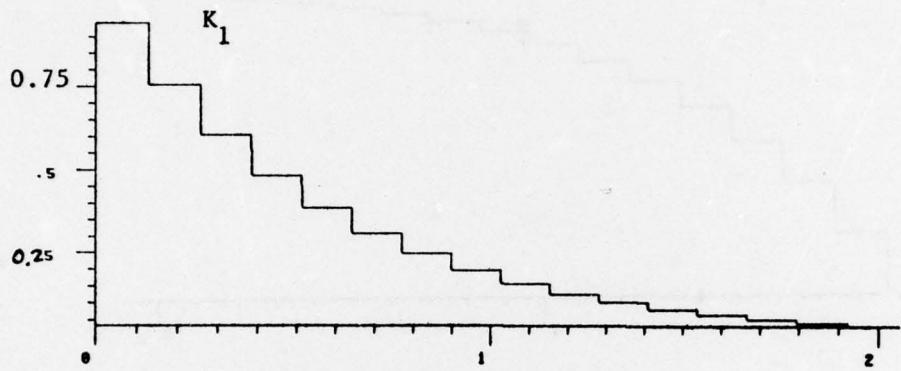


Figure 2(a). First reflection coefficient,  $K_1$ , of underlying AR model.

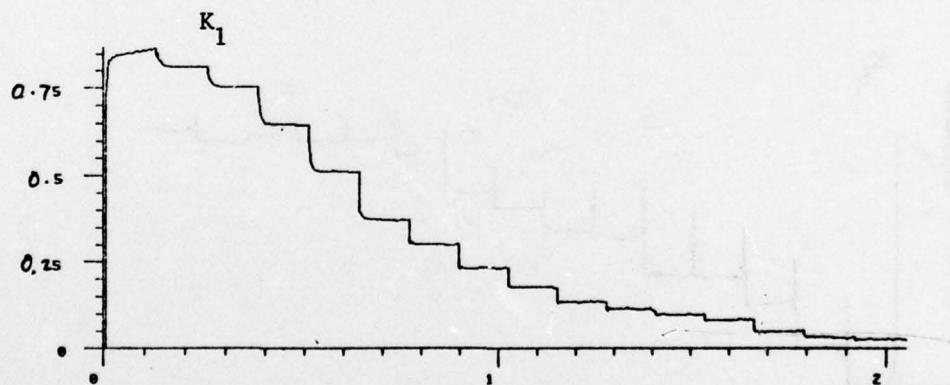


Figure 2(b). Estimated  $K_1$  using recursive ladder form.

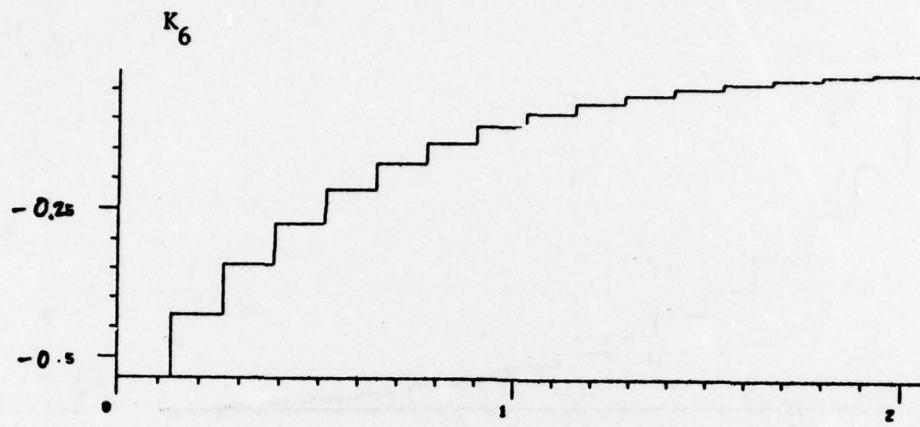


Figure 3(a). Sixth reflection coefficient,  $K_6$ , of underlying AR model.

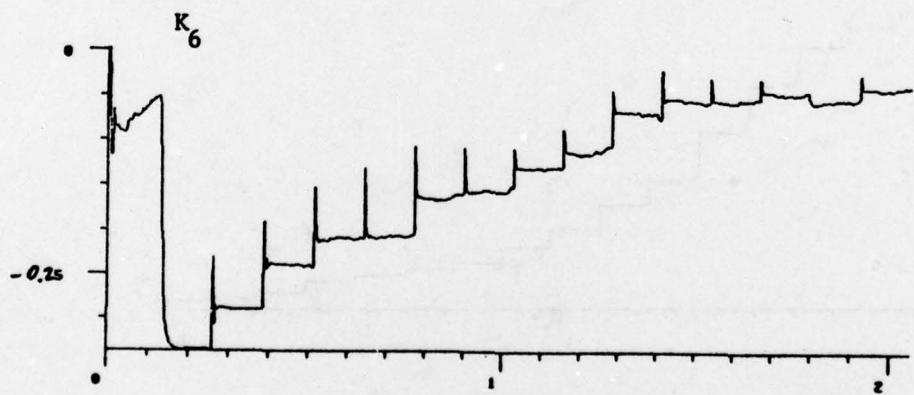


Figure 3(b). Estimated  $K_6$  using recursive ladder form.

IMPROVED SPECTRAL ESTIMATION FROM INCOMPLETE  
SAMPLED-DATA OBSERVATIONS\*

JAMES A. CADZOW

Department of Electrical Engineering  
Virginia Polytechnic Institute and State University  
Blacksburg, Virginia 24061

Abstract

In this paper, a particularly efficient procedure for achieving improved spectral estimations from incomplete observations of data sequences is presented. This problem has relavancy in such applications as Doppler radar signal processing where one needs to estimate Doppler frequency shifts based upon a very small number of radar returns. The essence of the method is that of appropriately estimating the unobserved (or missing) data and using the enlarged data base to generate the improved spectrum estimate. Clearly, the effectiveness of this method will be dependent on how well the missing data can be estimated. Empirical evidence accumulated to date indicates that this paper's procedure is effective as well as being computationally efficient.

I. Introduction

The spectral content of a continuous-time signal is of primary interest in a variety of interdisciplinary applications. In particular, given the signal  $x(t)$ , its spectrum is defined to be the magnitude of the associated Fourier transform

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (1)$$

The behavior of the spectrum,  $|X(\omega)|$ , as a function of  $\omega$  often provides information otherwise not readily apparent in the original time signal. It is with this in mind that a great deal of activity has been recently devoted to developing spectrum estimation techniques applicable to situations in which the signal  $x(t)$  is not completely observable. This lack of complete observability can result, for example, when one is able to observe the signal only over a finite time interval even though the signal is itself defined for all

---

\*This work was supported, in part, by the Signal Processing Section of the Surveillance Technology Branch of Rome Air Development Center through the Post Doctoral Program under Contract No. F30602-75-C-0122.

time, or, when only discrete-time samples of the signal are provided. In this paper, we shall be concerned with the task of estimating a signal's spectral behavior given only a "finite" set of sampled values of the signal. Spectral estimations based on these restrictions are of fundamental concern in such practical applications as typified by Doppler radar signal processing.

Let us first assume that the signal  $x(t)$  is uniformly sampled every  $\Delta$  seconds to generate the "infinite" length sequence  $\{x(n\Delta)\}$  in which  $n$  is the integer valued discrete-time variable. The spectrum of this sequence is formally obtained by evaluating its associated discrete-time Fourier transform as defined by

$$\bar{X}(\omega) = \sum_{n=-\infty}^{\infty} x(n\Delta) e^{-j\omega n\Delta} \quad (2)$$

where the overbar is used to distinguish this transform from the corresponding continuous-time Fourier transform (1). It is apparent that  $\bar{X}(\omega)$  is a periodic function of  $\omega$  with period  $2\pi/\Delta$ . Moreover, it is also readily established that the following relationship exists between the Fourier transforms (1) and (2)

$$\bar{X}(\omega) = \frac{1}{\Delta} \sum_{k=-\infty}^{\infty} X(\omega - 2\pi k/\Delta) \quad (3)$$

Although this relationship is true for any choice of the sampling time parameter  $\Delta$ , it is particularly meaningful when the continuous-time signal is band-limited in the sense that  $X(\omega) \equiv 0$  for  $|\omega| \geq \omega_1$ , and, when the sampling time is selected to satisfy the Nyquist criterion  $\Delta < \pi/\omega_1$ . In this important special case, it is clear from expression (3) that

$$X(\omega) = \Delta \bar{X}(\omega) \quad \text{for } |\omega| < \pi/\Delta \quad (4)$$

which implies that the continuous-time signal's Fourier transform can always be recovered from the associated discrete-time signal's Fourier transform under the stated conditions.

In any real-world application, however, it must be appreciated that one has available only a finite number of samples upon which to estimate the spectrum. Specifically, there will be available only the following partial observation (usually finite in number) of the underlying sequence

$$x(n\Delta) \quad \text{for } n \in \Lambda \quad (5)$$

where the "observation set"  $\Lambda$  consists of an incomplete integer set (i.e.,  $\Lambda \neq \{n: -\infty < n < \infty\}$ ). When this observation set consists of a contiguous set of integers (i.e.,  $\Lambda = \{n: n_1 \leq n \leq n_2\}$ ), then the resultant observed sequence is a standard truncated version of the underlying infinite length sequence. We shall not so restrict  $\Lambda$ , however, for we would then exclude such important situations as: (1) when data elements are missing, or, (2) when sequence interpolation is required.

To estimate the spectrum of the underlying infinite length sequence  $\{x(n\Delta)\}$  from partial observation (5), a natural procedure would be to evaluate the following partial Fourier transform

$$\bar{X}_p(\omega) = \sum_{n \in \Lambda} x(n\Delta) e^{-j\omega n\Delta} \quad (6)$$

It is clear that when the observation set  $\Lambda$  consists of all integers, then this partial Fourier transform and the underlying Fourier transform (2) are identical. Unfortunately, when  $\Lambda$  consists of only a moderate sized set of integers as is typical in many applications, this estimate is generally of poor quality. It is with this in mind that a variety of alternate spectrum estimation procedures have been recently developed. These include the essentially equivalent autoregressive and maximum entropy methods (e.g., see refs. [1]-[3]) and various extrapolation techniques (e.g., see refs. [4]-[7]). In this paper, we shall develop a procedure for estimating the behavior of the underlying sequence outside the observation set  $\Lambda$  with the objective of obtaining an improved spectral estimate from the resultant enlarged data base. The task to be then considered is given by the following

SEQUENCE RECONSTRUCTION PROBLEM: Let  $\{x(n\Delta)\}$  be a  $\Omega$  band-limited sequence in the sense that its Fourier transform  $X(\omega)$  as given by expression (2) is such that

$$\bar{X}(\omega) \equiv 0 \quad \text{for } \omega \notin \Omega \quad (7)^1$$

where  $\Omega$  is a subset of the frequency interval  $-\pi/\Delta < \omega < \pi/\Delta$  which has nonzero measure. Given the incomplete observation

$$x(n\Delta) \quad \text{for } n \in \Lambda \quad (8)$$

of the infinite length sequence  $\{x(n)\}$  where  $\Lambda$  is an incomplete set of integers, estimate values for the unobserved portion of the infinite length sequence (i.e., find  $x(p\Delta)$  for  $p \notin \Lambda$ ).

In formulating the reconstruction problem in this manner, we are able to simultaneously consider the apparently different special cases of low-pass, band-pass, and high-pass sequences.<sup>2</sup>

This sequence reconstruction problem is somewhat ill-posed in the sense that there exist an infinity of different  $\pi/\Delta$  band-limited continuous-time

---

<sup>1</sup>The condition  $\omega \notin \Omega$  is here meant to be all real numbers  $\omega$  in the interval  $(-\pi/\Delta, \pi/\Delta)$  not contained in  $\Omega$ .

<sup>2</sup>The sets  $\Omega$  which correspond to these special cases are  $\Omega_1 = \{\omega: |\omega| < \omega_1\}$ ,  $\Omega_2 = \{\omega: \omega_0 < |\omega| < \omega_1\}$ , and  $\Omega_3 = \{\omega: \omega_1 < |\omega| < \pi/\Delta\}$ , respectively.

signals which possess the same incomplete observation (8). To establish this conjecture, one need simply apply the well-known rule for reconstructing a band-limited signal from its uniformly sampled version  $\{x(n\Delta)\}$  where  $\Delta$  has been selected to satisfy the Nyquist criterion (e.g., see ref. [10], p. 29), that is

$$\begin{aligned}\hat{x}(t) &= \sum_{n \in \Lambda} x(n\Delta) \frac{\sin[\pi(t - n\Delta)/\Delta]}{\pi(t - n\Delta)/\Delta} + \sum_{n \notin \Lambda} x(n\Delta) \frac{\sin[\pi(t - n\Delta)/\Delta]}{\pi(t - n\Delta)/\Delta} \\ &= x_o(t) + x_u(t)\end{aligned}\quad (9)$$

We have suggestively decomposed this summation into two components which reflect the "observed" and "unobserved" elements of the underlying sampled sequence  $\{x(n\Delta)\}$ . Clearly, "any selection" for the unobserved samples (i.e.,  $x(n\Delta)$  for  $n \notin \Lambda$ ) will not effect the behavior of the reconstructed signal (9) at the observation times (i.e.,  $\hat{x}(n\Delta) = x(n\Delta)$  for  $n \in \Lambda$ ). Since there exists an infinity of such unobserved selections, we have proven the following lemma.

Lemma 1: There exists an infinity of different  $\pi/\Delta$  band-limited continuous-time signals which have the specified incomplete observed sampled values (8).

The set of signals alluded to in this lemma, in fact, is composed of the linear variety (i.e., a translated subspace) given by

$$V_1 = x_o + M$$

where  $x_o$  is given by the fixed observed first term on the right side of relationship (9) while  $M$  is the closed subspace spanned by the basis vectors  $\phi_n(t) = \Delta \sin[\pi(t - n\Delta)/\Delta]/\pi(t - n\Delta)$  for  $n \notin \Lambda$ .

The band-limited continuous-time signal which generated the observed samples (8) is then known to lie in linear variety  $V_1$ . If we were to uniformly sample each of the signals in  $V_1$ , there would result an associated linear variety of infinite length sequences. Generally, only a small subset of these sequences will be  $\Omega$  band-limited in the sense (7) with the overwhelming number not being band-limited at all (i.e., their Fourier transforms being zero only on zero measure subsets of the interval  $(-\pi/\Delta, \pi/\Delta)$ ). Our interest is clearly confined to those continuous-time signals in  $V_1$  which give rise to infinite length sequences which are  $\Omega$  band-limited. The proposed reconstruction problem will be well-posed only if there exists one such signal in  $V_1$ .

## II. Vector Space Formulation

It will be advisable to restate the reconstruction problem in a vector space setting so as to make use of the many powerful methods of linear

operator theory. In particular, we shall direct our attention to the set of sequences which possess finite energy. The set of finite energy sequences is known to constitute a Hilbert space (e.g., see ref. [8]) which is denoted by the symbol  $\ell_2$  and is specified by

$$\ell_2 = \{x: \langle x, x \rangle \text{ is finite}\} \quad (10)$$

where we have chosen to represent the sequence  $\{x(n)\}$  by the more compact notation  $x$ . Furthermore, the operator  $\langle \cdot, \cdot \rangle$  used in defining space  $\ell_2$  is the standard sequence inner product as defined by

$$\langle x_1, x_2 \rangle = \sum_{n=-\infty}^{\infty} x_1(n) x_2^*(n) \quad (11)$$

in which  $x_2^*(n)$  denotes the complex conjugate of  $x_2(n\Delta)$ . It can be shown that any sequence contained in Hilbert space  $\ell_2$  possesses a Fourier transform as given by relationship (2). Moreover, the generating sequence elements can be recovered from its associated Fourier transform by means of the inverse Fourier transform relationship

$$x(n\Delta) = \frac{\Delta}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} \bar{X}(\omega) e^{j\omega n\Delta} d\omega \quad (12)$$

Our primary interest in Hilbert space  $\ell_2$  will be concerned with the subset of sequences that are band-limited relative to a given nonzero measure frequency subset  $\Omega$  of the interval  $-\pi/\Delta < \omega < \pi/\Delta$ . This sequence subset will be denoted by  $B(\Omega)$  in which

$$B(\Omega) = \{x \in \ell_2: \bar{X}(\omega) \equiv 0 \quad \text{for } \omega \notin \Omega\} \quad (13)$$

where  $\bar{X}(\omega)$  denotes the Fourier transform of sequence  $x$ . It is a relatively simple matter to show that this subset is in fact a closed subspace of  $\ell_2$ . In addition, there exists a companion closed subspace of  $B(\Omega)$ , known as its orthogonal complement, which is defined by

$$B(\Omega)^\perp = \{x \in \ell_2: \bar{X}(\omega) \equiv 0 \quad \text{for } \omega \in \Omega\} \quad (14)$$

Using the discrete-time version of Parsevals' theorem, it immediately follows that these two closed linear subspaces are orthogonal (i.e.,  $\langle x_1, x_2 \rangle = 0$  for all  $x_1 \in B(\Omega)$  and  $x_2 \in B(\Omega)^\perp$ ).

Since the set  $B(\Omega)$  is a closed subspace of  $\ell_2$ , the following direct sum decomposition of Hilbert space  $\ell_2$  is evident

$$\ell_2 = B(\Omega) \oplus B(\Omega)^\perp$$

This implies that any sequence  $x \in \ell_2$  can be expressed uniquely as  $x = x_1 + x_2$  with  $x_1 \in B(\Omega)$  and  $x_2 \in B(\Omega)^\perp$ . For reasons which will be shortly made apparent, a procedure for effecting this decomposition will now be given. This is readily achieved by equivalently expressing the inverse Fourier transform relationship (12) as follows

$$\begin{aligned} x(n\Delta) &= \frac{\Delta}{2\pi} \int_{\omega \in \Omega} \bar{X}(\omega) e^{j\omega n\Delta} d\omega + \frac{\Delta}{2\pi} \int_{\omega \notin \Omega} \bar{X}(\omega) e^{jn\omega\Delta} d\omega \\ &= x_1(n\Delta) + x_2(n\Delta) \end{aligned} \quad (15)$$

where the first integral has been set equal to  $x_1(n\Delta)$  and the second to  $x_2(n\Delta)$ . Clearly, the sequences  $x_1$  and  $x_2$  so generated will be contained in the closed subspaces  $B(\Omega)$  and  $B(\Omega)^\perp$ , respectively, and the required decomposition has been made.

An examination of integral relationship (15) indicates that the required sequence decomposition can, in fact, be achieved by passing the sequence  $x$  through the ideal  $\Omega$  band-pass digital filter whose frequency transfer function is given by

$$H(\omega) = \begin{cases} 1 & \text{for } \omega \in \Omega \\ 0 & \text{for } \omega \notin \Omega \end{cases} \quad (16)$$

In particular, the response of this ideal filter to the input sequence  $\{x(n)\}$  has the Fourier transform expression  $H(\omega) \bar{X}(\omega)$  which is seen to be identically zero for  $\omega \notin \Omega$ . Thus, this response sequence must be contained in subspace  $B(\Omega)$ . With this in mind, let us then express the Fourier transform of the sequence  $x$  using the identity

$$\bar{X}(\omega) = H(\omega) \bar{X}(\omega) + [1 - H(\omega)] \bar{X}(\omega) \quad (17)$$

The sequence which corresponds to  $H(\omega) \bar{X}(\omega)$  is then identified with  $\{x_1(n)\}$  while that which corresponds to  $[1 - H(\omega)] \bar{X}(\omega)$  is clearly equal to  $\{x_2(n)\}$  since  $1 - H(\omega)$  is identically one for  $\omega \notin \Omega$  and zero for  $\omega \in \Omega$ . Relationship (17) thereby yields the desired decomposition in the frequency domain.

#### Observation Operator

To complete our vector space formulation, we shall now introduce two linear operators defined on Hilbert space  $\ell_2$ . The first operator will be appropriately referred to as the observation operator,  $L$ , which is characterized by

$$y = Lx \quad (18a)$$

where

$$y(n\Delta) = \begin{cases} x(n\Delta) & \text{for } n \in \Lambda \\ 0 & \text{otherwise} \end{cases} \quad (18b)$$

Here,  $\Lambda$ , is the finite observation set used in defining the sequence reconstruction problem. Clearly, the observation operator produces only a partial observation of the sequence  $x$  being operated upon and as such it is not an invertible operator. Namely, given the sequence  $y$ , it is not possible to uniquely recover the sequence  $x$  which generated  $y$  unless other constraints on sequence  $y$  are imposed.

#### Ideal Band-Pass Operator

The second operator corresponds to the previously mentioned ideal  $\Omega$  band-pass filter as characterized by transfer function (16). We shall put this filtering operation into the more compact operator relationship

$$y = Px \quad (19a)$$

where the elements of the sequences  $x$  and  $y$  are related by the convolution summation

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) \quad (19b)$$

The unit-impulse response sequence which characterizes this filter is simply the inverse Fourier transform of the ideal transfer function (16), that is

$$h(n\Delta) = \frac{\Delta}{2\pi} \int_{\omega \in \Omega} e^{jn\omega\Delta} d\omega \quad (19c)$$

It is interesting to note that the set of  $\Omega$  band-limited sequences in fact corresponds to those sequences in  $\ell_2$  which are eigen-sequences of operator  $P$  with corresponding eigenvalue one, that is

$$B(\Omega) = \{x \in \ell_2 : x = Px\}$$

It is now possible to reformulate the sequence reconstruction problem using the vector space concepts presented in this section. Namely,

SEQUENCE RECONSTRUCTION PROBLEM: Let the sequence  $x \in B(\Omega)$  which in turn requires that

$$x = Px \quad (20a)$$

where  $P$  is the ideal  $\Omega$  band-pass operator (19). Furthermore, let there be provided an incomplete observation of this band-limited sequence as specified by

$$y = Lx \quad (20b)$$

where  $L$  is the observation operator (18). From the incomplete observation sequence  $y$ , estimate values for the unobserved portion of sequence  $x$ .

The reconstruction problem as here formulated has the potential of being ill-posed due to the basic nature of the operator  $L$  and subspace  $B(\Omega)$ . This will be the case if the null space of operator  $L$  as specified by

$$N(L) = \{x \in B(\Omega) : Lx = 0\} \quad (21)$$

is nontrivial. We reach this conclusion by noting that the solution to operator relationship (20) must lie in the linear variety

$$V_2 = \bar{x} + N(L) \quad (22)$$

where  $\bar{x}$  is any sequence which will satisfy expression (20). Clearly, if  $N(L)$  contains more than the zero sequence, there will exist an infinity of different sequences in  $B(\Omega)$  which will satisfy relationship (20). Fortunately, this potentially damaging ill-posedness will not be present when the subspace  $N(L)$  contains only the zero sequence, or, when it is possible to impose further restrictions on the class of  $\Omega$  band-limited sequences to be considered (e.g., sinusoidal sequences) which has the effect of causing  $N(L)$  to contain only the zero sequence. In any case, the investigator must appreciate the potentially intrinsic ill-posed nature of the problem at hand when offering solutions to the sequence reconstruction problem.

We shall conclude this section by making an important characterization of the observation and ideal band-pass operators  $L$  and  $P$ , respectively. Namely, it is a relatively simple matter to show that they are each idempotent (i.e.,  $L^2 = L$  and  $P^2 = P$ ) and that their range and null spaces are orthogonal (i.e.,  $R(L) \perp N(L)$ , and,  $R(P) \perp N(P)$ ). As such, these two operators which characterize the sequence reconstruction problem are orthogonal projection operators. This characterization can play a most vital role in any attempt at finding solutions to relationship (20).

### III. Sequence Reconstruction Procedure

Recently, a signal reconstruction algorithm for obtaining a solution to a more general version of operator relationship (20) has been developed [9]. This algorithm is based upon the method of successive corrections and takes the form

$$x_n = x_{n-1} - PLx_{n-1} + P(Lx) \quad n=1,2,3,\dots \quad (23)$$

where we have suggestively expressed the observed signal  $y$  by its equivalent  $Lx$ . The initial approximation sequence  $x_0$  is required to lie in  $B(\Omega)$  and is typically selected to be the zero sequence. Using the fact that  $P$  and  $L$  are orthogonal projection operators, it has been shown that this reconstruction algorithm generates a sequence (of sequences)  $\{x_n\}$  which converges in the sense that (see ref. [9])

$$x_n \rightarrow x + u \quad (24)$$

The sequence  $u$  is equal to the orthogonal projection of the initial approximation error  $x_0 - x$  onto the subspace consisting of  $\Omega$  band-limited sequences which lie in the null space of operator  $L$ , that is

$$N(L) = \{x \in B(\Omega): Lx = \theta\}$$

When the sequence  $x_0 - x$  is nearly orthogonal to  $N(L)$ , it follows that  $u \approx \theta$  and the algorithm will converge to the desired result  $x$ . In general, this may not be the case and the generated sequence  $\{x(n\Delta)\}$  can converge to a sequence other than the original  $x$ . It has been empirically determined that in many applications, the desired condition  $x_0 - x$  being orthogonal to  $N(L)$  is nearly met thereby resulting in a desired convergence property. It is important to note, however, that the effectiveness of proposed algorithm (23) (or many other reconstruction procedures) will be dependent on the nature of the class of  $\Omega$  band-limited sequences to be considered.

Although reconstruction algorithm (23) will generate an approximation sequence which has guaranteed convergence, the convergence rate can be disappointingly slow. In recognition of this undesirable situation, a direct solution procedure has been recently developed which will yield the desired reconstruction [4]. Namely, one first finds the sequence  $z \in R(L)$  (i.e.,  $z(n) = 0$  for  $n \notin \Lambda$ ) which will satisfy the operator relationship

$$\text{Step 1: } LPz = Lx \quad (25)$$

Once this sequence has been found, the reconstructed sequence is obtained by operating on  $z$  by  $P$ , that is

$$\text{Step 2: } x + u = Pz \quad (26)$$

This two-step procedure will yield the reconstructed sequence so long as the solution  $z$  obtained from operator relationship (24) is such that

$$\lim_{k \rightarrow \infty} [LP]^k z = \theta \quad (27)$$

Any reasonably well-behaved sequence  $z$  will eventually be annihilated by an infinite number of sequential applications of the composite operator  $LP$ . Thus, requirement (27) offers no real restriction in terms of finding a solution  $z$ . This convenient reconstruction procedure has been found to be very effective in all examples treated to date and avoids the typically slow convergent rates so characteristic of signal reconstruction algorithms (e.g., [5] and [6]).

#### IV. Sequence Extrapolation and Interpolation

Two relevant reconstruction tasks which have received much attention in the literature are that of the extrapolation or interpolation of band-limited sequences. We shall consider the specific case in which the band-limit set  $\Omega$  is given by

$$\Omega = \{\omega: \omega_0 < |\omega| < \omega_1\} \quad (28)$$

where  $\omega_0$  and  $\omega_1$  are fixed numbers contained in the interval  $[0, \pi/\Delta]$ . The ideal unit-impulse response which corresponds to this  $\Omega$  band-limit set is given by expression (19c), that is

$$h(n\Delta) = \frac{\sin(\omega_1 n\Delta) - \sin(\omega_0 n\Delta)}{n\pi} \quad (29)$$

We shall now consider separately, the extrapolation and interpolation problems which correspond to this  $\Omega$  band-limit set.

##### Extrapolation

In the extrapolation task, the observation set will be taken to be the contiguous set of integers

$$\Lambda = \{M, M + 1, \dots, N - 1, N\} \quad (30)$$

where the integers  $M$  and  $N$  define the observation interval. With this selection of  $\Omega$  and  $\Lambda$ , operator relationship (25) becomes

$$\sum_{k \in \Lambda} \frac{\sin([n - k]\omega_1\Delta) - \sin([n - k]\omega_0\Delta)}{\pi(n - k)} z(k\Delta) = x(n\Delta) \quad \text{for } n \in \Lambda \quad (31)$$

where use of the fact that  $z(k\Delta) \equiv 0$  for  $k \notin \Lambda$  has been made. We next solve this consistent system of  $N + 1 - M$  linear equations for the  $N + 1 - M$  unknowns  $z(k\Delta)$  with  $k \in \Lambda$ . Once this solution has been determined, the desired extrapolation is achieved by incorporating relationship (26), that is

$$x(p\Delta) = \sum_{k \in \Lambda} \frac{\sin([p - k]\omega_1\Delta) - \sin((p - k)\omega_0\Delta)}{\pi(p - k)} z(k\Delta) \text{ for } p \notin \Lambda \quad (32)$$

In essence, this extrapolation requires the solution of the system of  $N + 1 - M$  linear equations (31) and then an evaluation of relationship (32) for each extrapolation element desired. Clearly, the computational requirements of this algorithm are minimal. The most time consuming aspect of this procedure is that of solving relationship (31) which requires on the order of  $(N + 1 - M)^3$  multiplications (or even fewer if more specialized procedures such as Levinson's method are used).

#### Extrapolation Example

We shall now illustrate this extrapolation procedure by considering the case where the parameter selections  $\omega_1 = 2\pi/25\Delta$  and  $\omega_0 = 0.8\omega_1$  are made. Furthermore, the underlying continuous-time signal  $x(t)$  to be extrapolated is given by

$$x(t) = \sin(0.99 \omega_1 t) + \sin(0.85 \omega_1 t)$$

The sequence which results when this signal is uniformly sampled is seen to be

$$x(n\Delta) = \sin(0.99 n\omega_1\Delta) + \sin(0.85 n\omega_1\Delta)$$

in which  $\omega_1\Delta = 2\pi/25$ . It is apparent that the spectrum of  $\{x(n\Delta)\}$  is entirely contained within the specified band-limit set  $\Omega = \{\omega: 0.8\omega_1 < |\omega| < \omega_1\}$ . Thus, the extrapolation procedure as characterized by relationships (31) and (32) is applicable. Finally, the observation set  $\Lambda$  is taken to be

$$\Lambda = \{n: -5 \leq n \leq 5\}$$

which corresponds to only 0.374 of one period of the lower frequency sinusoidal component constituting  $x(n\Delta)$ .

Using relationship (31) with the above selection of  $\omega_0$ ,  $\omega_1$ , and  $\Lambda$ , there will result a system of 11 Toeplitz equations to solve for the eleven unknowns  $z(n\Delta)$  for  $n \in \Lambda$ . Using a modified version of the Levinson algorithm, this system of equations is solved and that solution is inserted into the extrapolation expression (32) which is then evaluated for  $p = \pm 6, \pm 7, \dots, \pm 99$ . The results of this procedure are shown in Figure 1 in which the continuous line corresponds to the underlying continuous-time signal and the crosses outside the observation window denote the extrapolations while those within the window denote the eleven samples of  $x(t)$  used in the extrapolation procedure. Obviously, the extrapolation procedure has yielded virtually

exact results. If standard spectral estimation techniques such as the auto-regressive or maximum entropy procedures are applied to this extrapolated sequence of length 199, the resultant spectrum possesses very narrow peaks at radian frequencies  $0.99\omega_1$  and  $0.85\omega_1$ . This is to be contrasted with the spectrum estimates from the original unextrapolated sequence of length 11 which are not well defined at all. This behavior is demonstrated in Figure 2.

### Interpolation

For ease of presentation, the observation set  $\Lambda$  which corresponds to the interpolation reconstruction problem is taken to be

$$\Lambda = \{M, M + q, M + 2q, \dots, M + (N - M)q\} \quad (33)$$

where  $q$  is a positive integer greater than one and  $M$  and  $N$  are integers which determine the number of observations made (i.e.,  $N + 1 - M$ ). It is apparent from this observation set structure that we have available every  $q^{\text{th}}$  sample of the underlying sequence  $\{(x(n\Delta))\}$  commencing at discrete-time  $M$  and concluding at  $M + (N - M)q$ . With this choice of  $\Lambda$ , operator relationship (25) becomes

$$\sum_{k \in \Lambda} h([n - k]\Delta)z(k\Delta) = x(n\Delta) \quad \text{for } n \in \Lambda \quad (34)$$

where  $h(n\Delta)$  is specified by relationship (29). We now solve this consistent linear system of  $N + 1 - M$  equations in the  $N + 1 - M$  unknowns  $z(k\Delta)$  for  $k \in \Lambda$ . Finally, this solution is substituted into relationship (26) to obtain the desired reconstruction, that is

$$x(p\Delta) + u(p\Delta) = \sum_{k \in \Lambda} h([p - k]\Delta)z(k\Delta) \quad \text{for } p \notin \Lambda \quad (35)$$

Our interest in this "interpolation" problem is typically, though not necessarily, confined to evaluating relationship (35) for integers  $p$  which fill in the gaps of the observation set  $\Lambda$  (i.e.,  $p = M + 1, M + 2, \dots, M + q - 1, M + q + 1, \dots$ ). As in the extrapolation problem, the computational requirements of this interpolation procedure are modest. Namely, we solve the system of  $N + 1 - M$  linear equations (34) and then evaluate relationship (35) for each interpolation element desired.

### V. Conclusion

An efficient one-step procedure for reconstructing a band-limited sequence from an incomplete observation of that sequence has been presented. This method basically entails the solving of a system of consistent linear equations where the number of equations and unknowns equals the number of

sequence observations. The potential ill-posed nature of the underlying problem as well as procedures for circumventing this difficulty were discussed. The effectiveness of the proposed sequence reconstruction method was demonstrated by means of an example.

#### VI. References

- [1] Parzens, E., Multiple Time Series Modeling, Multivariate Analysis, Vol. 2, Edited by P. R. Krishnaiah, Academic Press, Inc., New York, 1969.
- [2] Burg, J. P., "Maximum Entropy Spectral Analysis," Ph.D. Dissertation, Stanford University, May 1975.
- [3] Van Den Bos, A., "Alternative Interpretation of Maximum Entropy Spectral Analysis," IEEE Trans. Information Theory, IT-17, pp. 493-494, July, 1971.
- [4] Cadzow, J. A., "An Extrapolation Procedure for Band-Limited Signals," submitted for publication.
- [5] Gerchberg, R. W., "Super Resolution through Error Energy Reduction," Optica Acta, Vol. 21, No. 9, pp. 709-720, 1974.
- [6] Papoulis, A., "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," IEEE Trans. on Circuits and Systems, Vol. CAS-22, No. 9, pp. 735-742, Sept. 1975.
- [7] Sabri, M. S., and W. Steenaart, "An Approach to Band-Limited Signal Extrapolation: The Extrapolation Matrix," IEEE Trans. on Circuits and Systems, Vol. CAS-25, No. 2, pp. 74-78, Feb. 1978.
- [8] Naylor, A. W., and G. R. Sell, Linear Operator Theory, Holt, Rinehart, and Winston, Inc., New York, 1971.
- [9] Cadzow, J. A., "Reconstruction of Signals from Their Linear Mapping Image," 1978 IEEE International Conference on Acoustics, Speech, and Signal Processing; Tulsa, Oklahoma, April 1978.
- [10] Oppenheim, A. V., and R. W. Schafer, Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

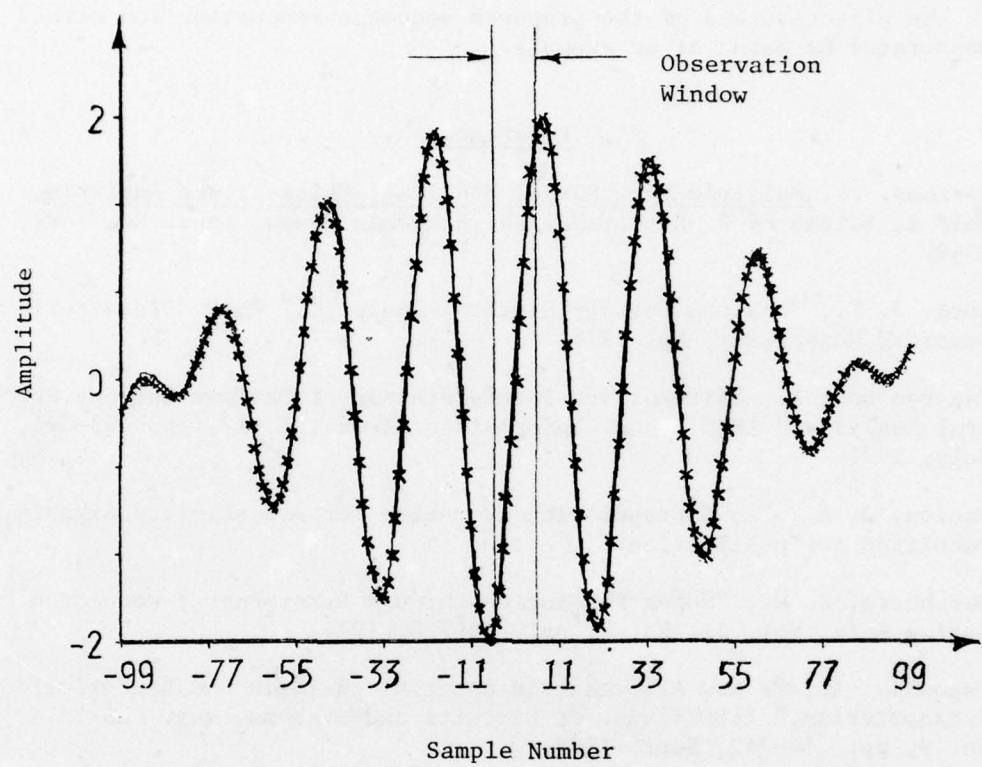


Fig. 1 Extrapolation of Time-Truncated Version  
of  $\sin[0.99\omega_1\Delta] + \sin[0.85\omega_1\Delta]$  Signal

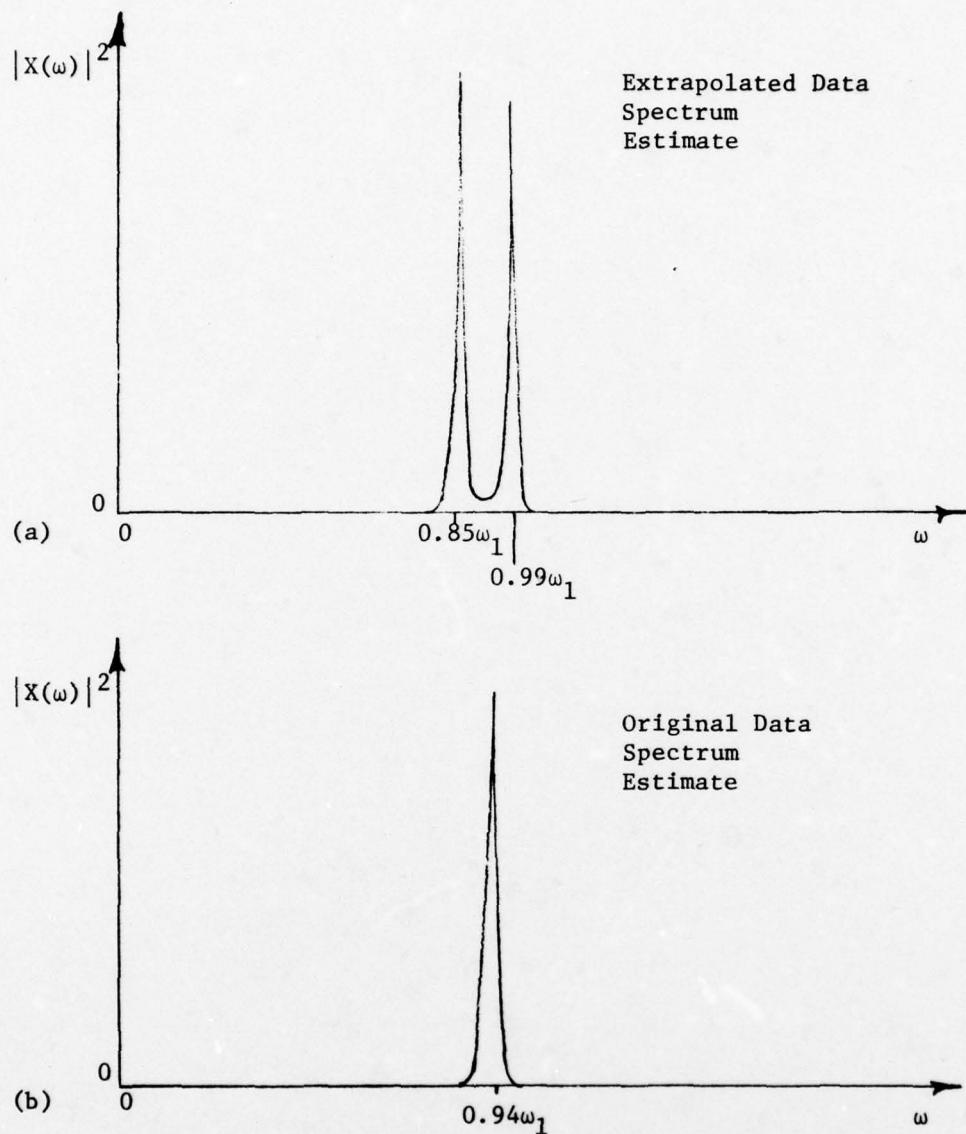


Fig. 2 Spectrum Estimation Plots: (a) extrapolated sequence of length 199 and a 15<sup>th</sup> order autoregressive model, (b) original sequence of length 11 and a 5<sup>th</sup> order autoregressive model.

# A REVIEW OF PRONY'S METHOD TECHNIQUES FOR PARAMETER ESTIMATION

MICHAEL L. VAN BLARICUM

Mission Research Corporation  
735 State Street, P. O. Drawer 719  
Santa Barbara, CA 93102

## Introduction

In analyzing and characterizing experimental or numerical electromagnetic response data one desires to extract parameters that can be related to physical characteristics of the system being studied. One obvious set of physically related parameters are the complex resonances of the system and their related coefficients. Indeed, spectral data usually lends itself to visually identifying these natural frequencies, however the damping constants cannot be obtained as easily. Similarly, temporal data generally allows one to visually determine the dominant natural frequency and if enough data is present, its damping constant.

Mains and Moffatt[1] introduced the concept of using complex natural resonances of a target as a basis for target recognition. They made use of the fact that a few natural resonances of a body are adequate to distinguish the body within a finite collection of bodies. Baum[2] developed the formalism known as the Singularity Expansion Method (SEM) which enables one to write any electromagnetic response of a system in an expansion of these complex resonances or poles and residues. Both of the above methods were based on obtaining the resonances from a set of equations which characterized the electromagnetic response of the body, much as a circuit theorist finds his resonances by solving a differential equation. The problem still exists of how one obtains these resonances from response data. In particular, how to obtain the parameters from transient response data from EMP simulators and transient radar ranges.

About four years ago Prony's algorithm[3] was once more dusted off and applied to this problem of information extraction from electromagnetics data. The first application at that time of Prony's method was to numerically generated transient current on a thin dipole. The results, which were reported at the 1974 USNC/URSI meeting[4] by this author gave a set of poles which compared very closely, if not exactly, with the first ten even modes previously calculated by Tesche[5]. As a result of this initial demonstration, several researchers began studying Prony's method to determine its utility for analyzing several kinds of transient data and to look for solutions to some

of the problems inherent in the process. In addition, Brittingham, Miller and Willows[6] demonstrated that a procedure parallel to the time domain Prony's method could be applied to frequency domain data.

Some of the initial questions which were asked about Prony's method and which are in part still being studied were:

1. Will Prony's method work if multiple poles are present?
2. How does one determine a priori the order of the system?  
That is, how many poles are contained in the response data?
3. What effects does noisy data have on the Prony algorithm?
4. How do we insure or know the accuracy of the poles returned?
5. What direct applications does the Prony procedure have?

These questions were all addressed to some extent in this author's dissertation[7]. It was found that Prony's method would work for the case of multiple poles without any change in the pole searching algorithm. Two methods were discovered by which the order of the system could be determined. These methods are the Householder orthogonalization procedure and the Eigenvalue method. Examples of these methods applied to clean and noisy data can be found in this author's thesis[7] and appeared in the recent special EMP issue of AP-S[8]. The applications of Prony's method seem to be never-ending. These range from analysis of electrocardiograms[9] to radar target recognition[10]. The most important and useful application made to date has been to the analysis of EMP simulator data. The problem of noise and Prony's method is a very complex one and will be discussed later in this paper.

The remaining part of this paper will present a very brief introduction to the mathematics involved in the Prony algorithm. The problem of noise and its effect on Prony's method will be discussed and several solution methods will be mentioned. Finally, some examples of the Prony's method will be presented.

It should be kept in mind that this is intended as a review paper and will necessarily be brief. Since most of the discussion here has been detailed elsewhere, these will be cited as thoroughly as possible.

#### The Prony Algorithm

The Prony algorithm was first published in 1795[3]. Since that time the method has been described by many authors. A good description can be found in Hildebrand[11]. The method is based on the fact that the system we are modeling can be represented by

$$R(t) = \sum_{i=1}^N A_i e^{S_i t} \quad (1)$$

where  $R(t)$  is the response, the  $S_i$  are the complex poles and the  $A_i$  are the corresponding residues. The  $S_i$  can also be written as  $S_i = \alpha_i + j\omega_i$ . The  $\alpha_i$  are thought of as the damping constants, and the  $\omega_i$  are the natural frequencies in radians per second. Since, in practice, the measured data will appear as a set of discrete data, Equation (1) is rewritten as

$$R(t_n) = R_n = \sum_{i=1}^N A_i e^{S_i n \Delta t}, \quad n = 0, 1, \dots, M-1 \quad (2)$$

where  $\Delta t$  is the time step size and  $M$  is the total number of samples taken. The set of Equations (2) is seen to be  $M$  nonlinear equations in  $2N$  unknowns. If  $M$  is equal to or greater than  $2N$ , and if all  $\Delta t$  are equal, then this nonlinear set of equations can be solved using the Prony algorithm.

Prony recognized that the  $R_n$  in (2) must satisfy a difference equation of order  $N$  which may be written as

$$\sum_{p=0}^N \alpha_p R_{p+k} = 0, \quad k = 0, 1, \dots, \gamma-1, \quad (3)$$

where  $\gamma$  is the value of  $M - N$ . The roots  $Z_i$  of the algebraic equation

$$\sum_{p=0}^N \alpha_p Z^p = 0 \quad (4)$$

define the natural frequencies through

$$Z_i = e^{\frac{S_i \Delta t}{\gamma}}, \quad i = 1, 2, \dots, N. \quad (5)$$

If in Equation (3)  $\alpha_N$  is defined equal to 1, then the remaining  $\alpha_p$ 's may be obtained by solving the equation

$$\sum_{p=0}^{N-1} \alpha_p R_{p+k} = -R_{N+k}. \quad (6)$$

If  $2N$  data samples are used, the Equation (6) can be solved exactly for the  $\alpha$ 's. If more than  $2N$  samples are desired, then one can use a least-squares

type fit to (6). Once the  $\alpha_p$  have been found then the roots,  $Z_i = \exp(S_i \Delta t)$ , of Equation (4) can simply be found and the poles are then obtained as

$$S_i = \frac{\ln Z_i}{\Delta t} \quad (7)$$

It is a simple procedure to obtain the residues,  $A_i$ , by solving the matrix equation embodied in Equation (2) once the  $S_i$  are known.

#### The Noise Problem

The problem of Prony's method and noisy data became apparent the first time noisy data was analyzed with the technique. This author's dissertation [7] and two recent papers in the special EMP AP-S[8][12] transactions discuss some of the typical effects which noise has on the results of the method. Noise in the data has a tendency to give totally lousy results if some attempt is not made to correct for its existence. It is likely that noise will perturb the extracted poles in such a way that they are not at all similar to the true poles. Noise almost always makes the damping constant  $\alpha$  too large. While the effects of the noise are well known and documented, the cause of those effects are not well understood. Don Dudley[13][14] has done an excellent study of some of the reasons behind the noise problems. Obviously it is necessary to understand the cause of these problems before we can apply any rational procedures for correcting them. Following is a brief outline of what appears to be the major effect of the noise.

The key step to Prony's method is the solution for the coefficients  $\alpha_p$  in the difference Equation (6) repeated here for convenience;

$$\sum_{p=0}^{N-1} \alpha_p R_{p+k} = -R_{N+k} \quad , \quad k = 0, 1, \dots, \gamma-1 \quad . \quad (6)$$

However, in the solution for the  $\alpha_p$  we do not know the  $R_{p+k}$  exactly. The measured data  $Y_k$  can be expressed as

$$Y_k = R_k + e_k \quad (8)$$

where  $e_k$  is the error in the  $k$ th sample. Hence, Equation (6) can be rewritten as

$$\sum_{p=0}^{N-1} \alpha_p Y_{p+k} = -Y_{N+k} + W_k \quad , \quad k = 0, 1, \dots, \gamma-1 \quad . \quad (9)$$

where

$$W_k = \sum_{p=0}^N \alpha_p e_{p+k} \quad . \quad (10)$$

The  $w_k$  are the residuals. In solving Equation (9) in a least squared error sense, the  $w_k$  are minimized with respect to the  $\alpha_p$  as

$$\frac{\partial}{\partial \alpha_m} \sum_{k=0}^{\gamma-1} w_k^2 = 0 , \quad m = 0, 1, \quad (11)$$

which gives the normal equations of least squares as

$$\sum_{p=0}^{N-1} \alpha_p \sum_{k=0}^{\gamma-1} Y_{p+k} Y_{m+k} = - \sum_{k=0}^{\gamma-1} Y_{N+k} Y_{m+k} , \quad m = 0, \dots, N-1 . \quad (12)$$

If the  $w_k$  are independent and equally distributed with zero mean, then as  $\gamma \rightarrow \infty$  the  $\alpha_p$  converge to the true parameters. However, Equation (10) clearly shows that the  $w_k$  are not independent but are correlated, which results in biased estimates of the  $\alpha_p$ . Dudley[14] has shown that as the value of  $\gamma$  increases the  $\alpha_p$  do converge but converge to the wrong answer. It is this biased answer that causes the damping constants to be too large. However, why they consistently come out too large as opposed to too small is unknown. This biasing is probably the major cause for all the strange results which are known to occur in Prony's method when noise is present.

At present there appears to be as many methods for correcting for the noise problem as there are researchers studying the Prony algorithm. These methods include prefiltering or averaging the data, internally whitening the residuals, and multiple processing with statistical analysis of the resulting poles. The more successful methods will be briefly discussed here, but the reader is encouraged to study the original works to get a complete understanding of the procedures.

If the analyst is fortunate enough to have sufficient data he can apply averaging or smoothing procedures to the data to lower the standard deviation of the error. If  $N$  samples are averaged, then the new average data point is obviously

$$\bar{Y} = \sum_{i=1}^N \frac{Y_i}{N}$$

and the new standard deviation of this point is

$$\bar{\sigma} = \left[ \sum_{i=1}^N \frac{\sigma_i^2}{N} \right]^{1/2} \quad (13)$$

This says that in order to get an order of magnitude decrease in  $\sigma$ ,  $N$  must equal 100. Hence, large amounts of data are needed to get a significant

decrease in the noise level by averaging.

Measurement noise seems to manifest itself in Prony's method as high frequency components in the data. Hence, Prony's method needs to find many high frequency poles outside the band of the required results. This suggests operating on the original data with a low pass filter to remove the high frequency components. This basically is equivalent to averaging the data but is implemented in a different fashion. One method of filtering which has been used by Cordaro[15] with fair success is the two-pole Butterworth digital filter. It is this author's opinion that much work still needs to be done to find the optimum preprocessing or filtering techniques to use with Prony's method. It will likely turn out that depending on the type of data being studied, different processing schemes will have to be used.

Even once filtering or averaging are used, there will still be noise in the data. Hence, the residuals  $W_k$  of Equation (10) will still be correlated and the estimates of the difference equation coefficients  $\alpha_p$  will be biased. Two methods appear to be useful in correcting the problem of the biased estimates. These methods are the repeated least squares method and the iterative generalized least squares method. Space does not allow explanations of these methods. The reader is referred to excellent reports by Dudley[13][14] and Cordaro[13] which explain the details of these methods and show examples of the results. The state-of-the-art of these procedures applied to Prony's method warrants more research, but the early results bode optimistic results.

If large amounts of data are present, Prony's method can be applied to several windows in the same data record. If  $M$  windows are used, each resulting in  $N$  poles, then the results will be  $M$  by  $N$  poles. These poles can be correlated to see which are true system poles and which are poles due to the noise. The assumption is that the poles due to the noise will shift wildly from window to window, while the true poles will stay essentially unchanged. Once the true poles are determined by looking for clumps of  $M$  poles, then the mean value can be found by averaging. This technique has been used fairly successfully by Hudson and Lager[16] and by this author[17]. There appears to be one basic flaw in this technique, however. That is, if there is noise in the data which biases the estimates of  $\alpha_p$ , then the true poles will be biased. Hence, the true poles will probably clump around the wrong result. This aberration has not appeared to be a problem in previous tests and needs to be studied in the future.

#### Examples of Prony's Application

Three previously published examples of the use of Prony's method will be presented in this section. These examples were chosen because of some of the insight which they give into the various attributes of the method. Two sets of noisy data are presented. It is not the intention here to hide the problems of Prony's method by showing only good examples; rather, it is hoped

that by showing some of the good, encouraging results more research will be directed into trying to solve some of the problems that hinder the method.

Example 1—From Reference [7]

A 1.0 m dipole with a half-length-to-radius ratio of 100 was numerically modeled and excited with a broadside incident Gaussian pulse. The backscattered electric field was calculated as a function of time for uniform resistive loading of 0, 125, 250, 500, 750 and 1683 ohms per meter. These fields are shown in Figure 1. The loading of 1683 Ω/m was chosen because Tesche[17] calculated that at this value the dipole becomes critically damped, giving a double pole on the negative real axis. The trajectories of the first seven even poles extracted using Prony's method are shown in Figure 2. Note that for the value of 1683 Ω/m the first pole has split and moved toward the origin and toward infinity, which indicates that this value of loading gives an overdamped situation. This does not imply that Tesche's value is wrong but points out the differences in our numerical models for the dipole.

This trajectory plot is very illuminating in that it shows the effect of resistance on an antenna in terms of the natural resonances. This example also shows that accurate poles can be obtained from transient data that does not look like damped sinusoidal data. This example shows the ideal kind of results which one would hope to get from Prony's method. The results not only give useful parameters for the waveforms, they also tell us something about the physical characteristics of the antenna.

Example 2—From References [17] and [18]

In this example a pulse driving function, Figure 3a, was used to illuminate a dipole in a corner reflector and the terminal voltage of the antenna was measured as a function of time, Figure 3b. The measurements were taken on the HDL transient antenna range[19]. Prony's method was used to fit the driving function and the response functions with the results shown in Figures 3a and 3b, respectively. The closed form expression for the transfer function of the antenna was obtained analytically by evaluating

$$H(s) = \frac{\sum_{i=1}^N \frac{A_i}{s-s_i}}{\sum_{j=1}^M \frac{B_j}{s-s_j}} \quad (14)$$

where  $A_i$  and  $s_i$  are the residues and poles of the response voltage and  $B_j$  and  $s_j$  are the residues and poles of the driving waveform. The inverse Laplace transform of Equation (14) was taken to give the impulse response

$H(t)$  which is plotted in Figure 3c. Also in Figure 3c is the impulse response obtained by taking a conventional FFT of both the driving functions and the response functions, dividing the latter by the former, and then Fourier transforming back to the time domain.

This example shows that Prony's method can be used to obtain a closed form analytical expression of the transfer function of a structure from experimental data. The poles in the resulting transfer function should then be the true poles of the system. The poles are not plotted here because there is not previous data to compare them with. In this example Prony's method was used basically as a curve fitting scheme to give analytical expressions for the transient response of the measured data. Since the curve fit was in terms of complex exponentials the analytical Laplace transforms and deconvolution could be performed as in Equation (14). If a constrained Prony's method had been used, then the numerator poles of Equation (14) would contain the poles of the denominator and division would be easy. A constrained Prony's method has been developed by this author but has not yet been implemented.

Example 3—From Reference [20]

This example is taken from an MRC report [20] which contains several similar examples and should be referred to for more detail. The data used is the transient waveform measured by a D sensor on the central body of an electrical mockup of the FLTSATCOM satellite. The mockup was excited by capacitive drive current injection to determine the electrical characteristics of a satellite when excited by SGEMP. Figure 4a shows the actual D sensor data which was used with the Prony algorithm to produce the curve fit shown in Figure 4b. From this Prony fit the data was extrapolated to a later time, Figure 4c, than was available in the original measurement. Note the low frequency ringing which is now apparent but which did not rear its head in the original data of Figure 4a. Since the data was D data, the charge density on the body could be obtained by integrating the analytical expression for the response. The result of this integration is shown in Figure 4d. Finally, the Laplace transform of the D data obtained from the pole data is shown in Figures 4e and 4f. Note the dominate resonance which appears at about 10 megahertz.

This example shows that Prony's method can be applied to raw data to obtain a late time extrapolation of the response not obtainable from the measurement. It is also possible to perform an analytical integration on the data and analytical Laplace transform to obtain the spectral characteristics of the response. Note that because of the truncated data record it would have been very difficult to obtain an FFT of the data.

### Conclusions

This paper has briefly given an historical survey of the application of Prony's method to electromagnetics response data. The method itself was presented and some of the numerical difficulties which are encountered when noise is present were discussed with suggestions for remedying these problems. Finally, three examples of the application of Prony's method to data for information extraction were presented. This paper has actually only skimmed the surface of the method and its application and problems and is intended as an introduction to the method. The reader is encouraged to study the references listed.

It is the author's opinion that Prony's method is a very powerful tool for data analysis and parameter estimation when skillfully employed. This has been shown convincingly in the previous three examples. There are still many unanswered questions which need to be studied, many of which relate to the noise problems indicated in The Noise Problem section. More answers about the method have been uncovered in trudging ahead and using it than have been answered by studying the problem and waiting to apply it after all the problems have disappeared. The games of parameter estimation and data analysis are not easy ones, and we must make use of all available tools and learn to use the tools through application and experimentation.

### References

1. R. H. Mains and D. L. Moffatt, "Complex Natural Resonances of an Object in Detection and Discrimination," TR-3424-1, Ohio State University, Electroscience Laboratory, Columbus, Ohio, June 1974.
2. C. E. Baum, "On the Singularity Expansion Method for the Solution of Electromagnetic Interaction Problems," Interaction Note 88, December 11, 1971.
3. R. Prony, "Essai Expérimental et Analytique sur les Lois de la Dilatabilité de Fluides Elastiques et sur Celles de la Force Expansive de la Vapeur de L'Alkool, A Différentes Températures," J. L'Ecole Polytech, (Paris), Vol. 1, No. 2, pp. 24-76, 1795.
4. R. Mittra and M. L. Van Blaricum, "A Novel Technique for Extracting the SEM Poles and Residues of a System Directly from its Transient Response," 1974 Annual URSI Meeting, University of Colorado, Boulder, Colorado, October 14-17, 1974.
5. F. M. Tesche, "On the Analysis of Scattering and Antenna Problems using the Singularity Expansion Technique," IEEE Trans. Antennas Propagat., Vol. AP-21, No. 1, p. 53, January 1973.

6. J. N. Brittingham, E. K. Miller, J. L. Willows, "The Derivation of Simple Poles in a Transfer Function from Real Frequency Information," Lawrence Livermore Laboratory, UCRL-52050, April 6, 1976.
7. M. L. Van Blaricum, "Technique for Extracting the Complex Resonances of a System Directly from its Transient Response," Ph.D. Dissertation, Department of Electrical Engineering, University of Illinois, December 1975; also AFWL Interaction Note #301.
8. M. L. Van Blaricum and R. Mittra, "Problems and Solutions Associated with Prony's Method for Processing Transient Data," IEEE Trans. on Ant. and Prop., January 1978, pp. 174-182.
9. T. Y. Young and W. H. Higgins, "On the Representation of Electrocardiograms," IEEE Trans. Bio-Med. Electronics, Vol. BME-10, No. 3, pp. 86-95, July 1963.
10. L. W. Pearson, M. L. Van Blaricum and R. Mittra, "A New Method for Radar Target Recognition Based on the Singularity Expansion for the Target," 1975 IEEE International Radar Conference, Arlington, Virginia, pp. 452-456, April 21-23, 1975.
11. F. B. Hildebrand, Introduction to Numerical Analysis. New York: McGraw-Hill, 1956, pp. 378-382.
12. A. J. Poggio, M. L. Van Blaricum, E. K. Miller and R. Mittra, "Evaluation of a Processing Technique for Transient Data," IEEE Trans. on Ant. and Prop., pp. 165-173, January 1978.
13. D. G. Dudley, "Parametric Modeling of Transient Electromagnetic Systems," Radio Science (to appear).
14. D. G. Dudley, "Fitting Noisy Data with a Complex Exponential Series," Lawrence Livermore Laboratory, UCRL-52242, March 7, 1977.
15. T. Cordaro, "A Note on Representing a Transient Waveform by a Sum of Complex Exponentials," AFWL-TR-77-46, Air Force Weapons Laboratory, July 1977.
16. H. G. Hudson and D. L. Lager, "Observations on the Operation of the SEMPEX Code," Lawrence Livermore Laboratory, UCID-17440, September 22, 1976.
17. M. L. Van Blaricum, "An Analysis of Existing Prony's Method Techniques," 1977 Spring Fulmen Meeting, March 21, 1977.

18. M. L. Van Blaricum and D. Schaubert, "An Experimental Transient Transfer Function Via Prony's Method," 1976 USNC/URSI Meeting, University of Massachusetts, Amherst, Mass., 10-15 October 1976.
19. D. H. Schaubert, "Measurement of Impulse Response of Communications Antennas," HDL-TR-1832, November 1977.
20. M. L. Van Blaricum, "The Use of Prony's Method as an Analysis Tool Applied to the Electrical Excitation of FLTSATCOM," Mission Research Corporation Report, MRC-R-301, January 1977.

Figures

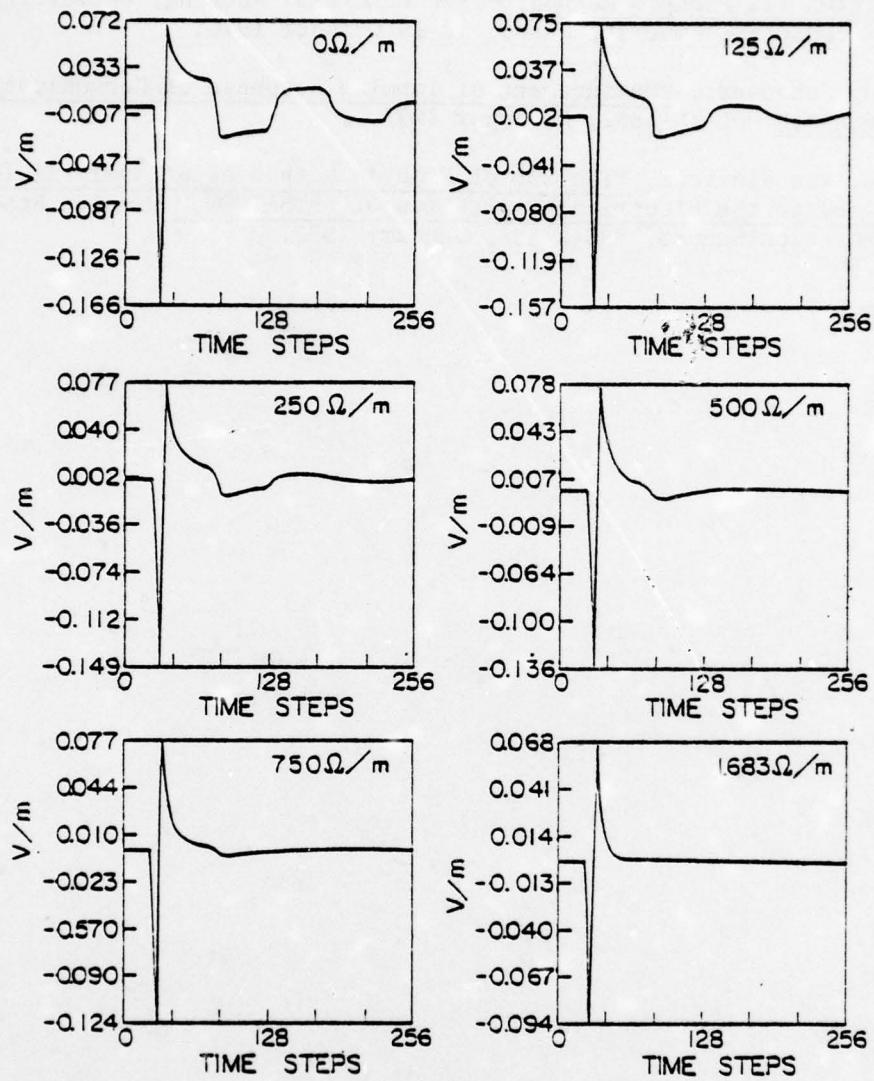


Figure 1. Backscattered fields for a 1 m dipole with resistive loading.  $\Delta t = 6.9444 \times 10^{-11}$  seconds

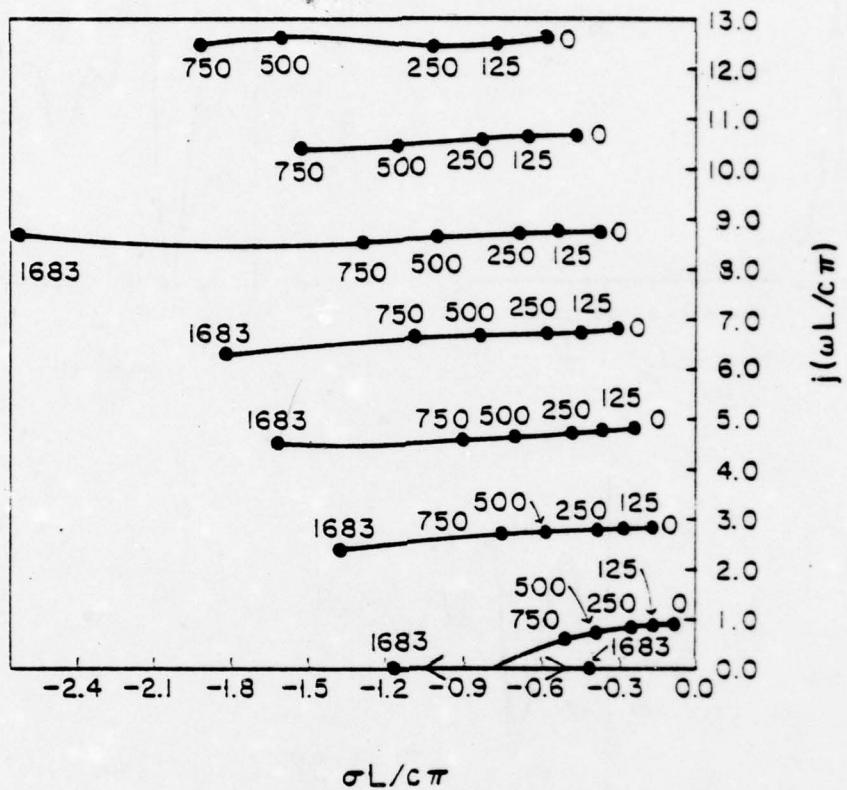


Figure 2. Trajectory of the poles for the uniformly resistive loaded 1.0 m dipole. Note the poles,  $S_i = \sigma + j\omega$ , are normalized by  $L = 1.0$  meter,  $c = 3 \times 10^8$  m/s, and  $\pi$ .

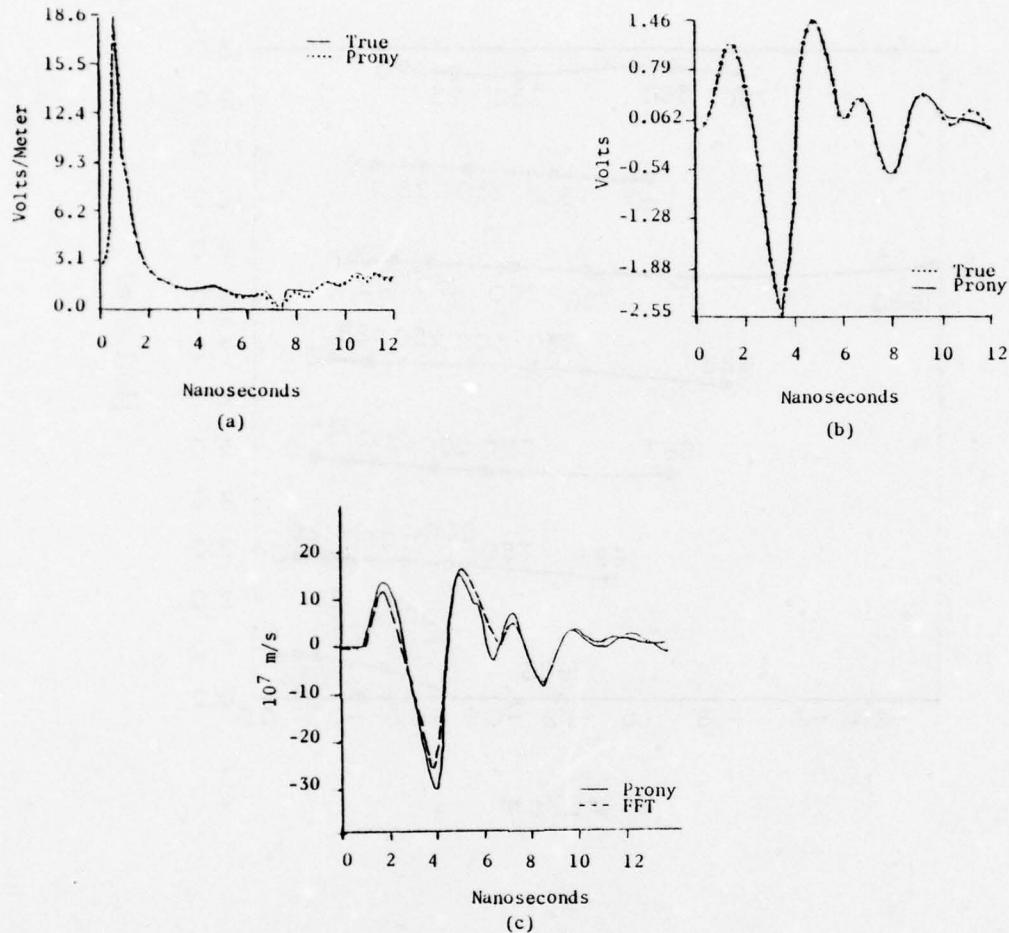


Figure 3. Data from the HDL transient antenna range for a dipole in a corner reflector. The data was operated on by Prony's method.

- (a) Original driving waveform and Prony fit.
- (b) Original measured response voltage and Prony fit.
- (c) Calculated impulse response of the antenna.

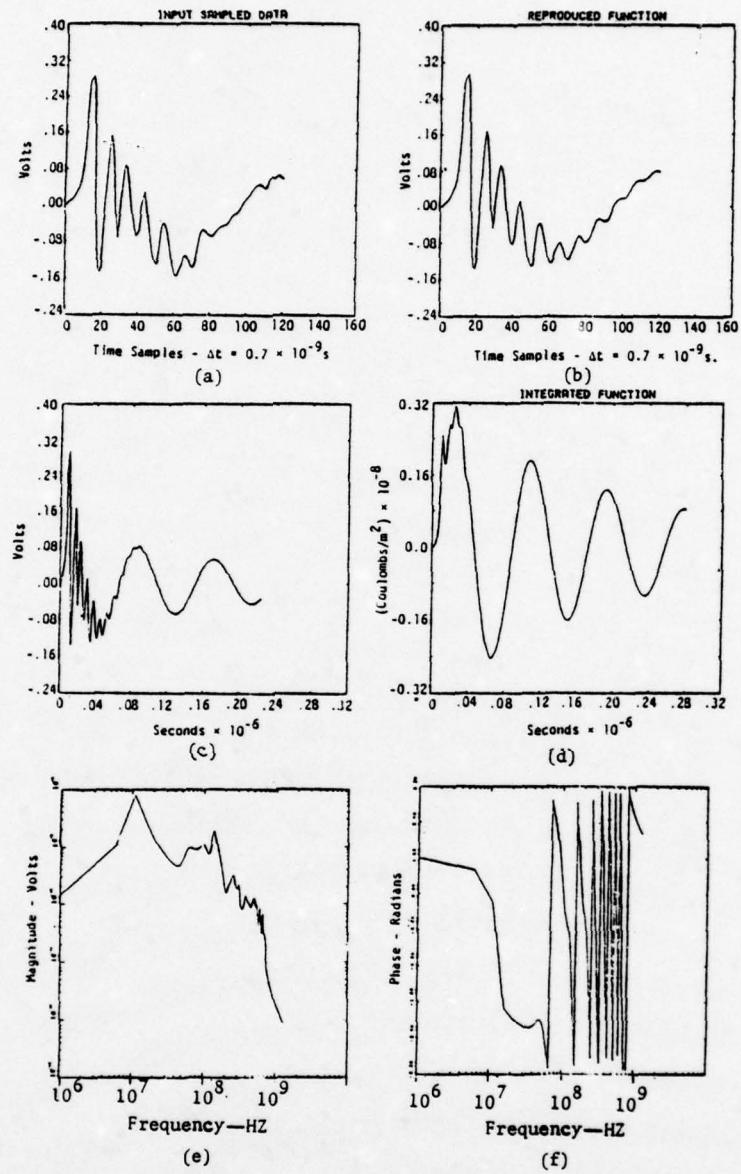


Figure 4. (a) Data from a FLTSATCOM electrical test; (b) operated on by Prony's method to obtain a curve fit; (c) an extrapolation; (d) analytical integration; and, (e) & (f) the Laplace transform.

## SYSTEM IDENTIFICATION BY USE OF PENCIL-OF-FUNCTIONS

V. K. JAIN

Department of Electrical Engineering  
University of South Florida  
Tampa, Florida 33620

D. D. WEINER            J. NEBAT  
Department of Electrical Engineering  
Syracuse University  
Syracuse, New York 13210

T. K. SARKAR

Department of Electrical Engineering  
Rochester Institute of Technology  
Rochester, New York 14623

### Abstract

Input-output responses of a network can be integrated to yield a family of signals, called measurement signals. Application of the pencil-of-functions theorem to this family yields, in a closed form, the identified parameters of the network function. This method of black-box modeling is suitable for systems where the responses can be integrated in real time, or recorded in digital form for off-line processing.

### INTRODUCTION

Determining the model of a network from its observed input-output responses represents the inverse of the analysis problem. Interest in this arises from the frequent need for a relatively simple mathematical description of the system so that behavior for other anticipated inputs may be predicted up to acceptable accuracies. Like the analysis problem, there are several approaches available in the literature for the inverse, or, as it is often called, the "identification" problem [1]. To name a few, a) Prony's method [2], b) gradient methods, such as Newton [3] and quasi-linearization [4], c) least-squares and generalized least-squares methods [5],[6], d) maximum-likelihood methods [7],[8], etc.

All of the methods stated above possess certain advantages and, as may be expected, certain disadvantages peculiar to each particular method.

Stated very broadly, sensitivity to noise, slow convergence to the solution, and excessive computational complexity are some of the possible disadvantages. The purpose of the present paper is to describe in a simple way the identification method developed in reference [9]. The method offers the advantages of mathematical simplicity, closed-form solution to the problem, which is optimal in the generalized least-squares sense and suboptimal in the strict least-squares sense, and relative robustness of the technique to noise. The disadvantage of the method is that unlike the maximum likelihood method, the variances of additive noise, when present, must be known a priori (since they are not estimated in the present method) in order that unbiased parameter estimates may be computed.

As stated above, we now set out to describe the pencil-of-functions method, without any proofs, and to illustrate it with some examples. Discrete-time signals are chosen for the presentation here, although such signals must often be obtained by sampling a continuous-time system.

#### PENCIL OF FUNCTIONS METHOD

##### Identification Problem

Given the input-output observations

$$\{u(k)\}, \{y(k)\}, \quad k=0,1,\dots,K \quad (1)$$

arising from a physical system believed to be linear, finite order, it is desired to find a system model

$$H(z) = \frac{b_1 z^{-1} + \dots + b_n z^{-n}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \quad (2)$$

$$= \sum_{i=0}^n \frac{d_i z^{-1}}{1 - c_i z^{-1}} \quad (3)$$

which best fits the observations, in some sense. A solution can be obtained by use of the pencil-of-functions theorem as discussed below.

For convenience denote sequences  $\{u(k)\}$  and  $\{y(k)\}$  simply as  $u$  and  $y$ , respectively. Also, denote the inner-product of two sequences as

$$x \cdot y \stackrel{\text{def}}{=} \sum_{k=0}^K x(k) y(k) \quad (4)$$

### Measurement Sequences

From the given sequences  $y$  and  $u$  we form the following set of sequences, called measurement sequences:

$$\begin{aligned} y_1(k) &= y(k) \\ y_2(k) &= y_1(0)+y_1(1)+\dots+y_1(k) \\ &\vdots \quad \vdots \quad \vdots \quad \cdots \quad \vdots \\ y_{n+1}(k) &= y_n(0)+y_n(1)+\dots+y_n(k) \end{aligned} \quad (5)$$

---


$$\begin{aligned} u_1(k) &= u(k-1) \\ u_2(k) &= u_1(0)+u_1(1)+\dots+u_1(k) \\ &\vdots \quad \vdots \quad \vdots \quad \cdots \quad \vdots \\ u_{n+1}(k) &= u_n(0)+u_n(1)+\dots+u_n(k) \end{aligned} \quad (6)$$

where  $n$  is the order of the model desired. That is,  $n$  is the degree of the network function  $H$ .

Note that these sequences represent repeated discrete integrations of the observed signals  $y(k)$  and  $u(k)$ , respectively. A schematic representation is given in Fig. 1 where we assume the additive noise to be zero, so that  $x(k)=y(k)$  and  $v(k)=u(k)$ . Correction for noise is discussed in [9].

### Gram Matrix

Next form the following inner-product matrix

$$F = \begin{bmatrix} y_1 \cdot y_1 & \dots & y_1 \cdot y_N & y_1 \cdot u_2 & \dots & y_1 \cdot u_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_N \cdot y_1 & \dots & y_N \cdot y_N & y_N \cdot u_2 & \dots & y_N \cdot u_N \\ u_2 \cdot y_1 & \dots & u_2 \cdot y_N & u_2 \cdot u_2 & \dots & u_2 \cdot u_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ u_N \cdot y_1 & \dots & u_N \cdot y_N & u_N \cdot u_2 & \dots & u_N \cdot u_N \end{bmatrix} \quad (7)$$

where we have used the notation  $N=n+1$  for convenience. This  $(N+n) \times (N+n)$  dimensional matrix is the Gram matrix [10] of the  $(N+n)$  dimensional vector sequence

$$\{\underline{f}(k)\}, \quad k=0, 1, \dots, K \quad (8)$$

where

$$\underline{f}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_N(k) \\ u_2(k) \\ \vdots \\ u_N(k) \end{bmatrix} \quad (9)$$

To state this observation formally, we have

$$F = \sum_{k=0}^K \underline{f}(k) \underline{f}^T(k) \quad (10)$$

### Diagonal Cofactors

Denote the diagonal cofactors of  $F$  as  $D_i$ :

$$D_i = i, i \text{ cofactor of } F \quad (11)$$

Recall that the  $i, i$  cofactor of a square matrix is the determinant of the matrix after deleting the  $i$ th row and the  $i$ th column.

### Parameter of the Network Function

The parameters of the network function are given by the square-roots of  $D_i$  up to a multiplicative constant. That is

$$\left[ \sum_{i=1}^N \sqrt{D_i} (1 - z^{-1})^{i-1} \right] Y(z) = \left[ \sum_{i=1}^n \sqrt{D_{N+i}} z^{-1} (1 - z^{-1})^{i-1} \right] U(z) \quad (12)$$

which can be normalized, by dividing by  $D = \sqrt{D_1} + \dots + \sqrt{D_N}$ , so that the leading coefficient becomes unity. Clearly the computed transfer function becomes

$$H(z) = \frac{z^{-1} \left[ \sum_{i=1}^n \sqrt{D_{N+i}} (1 - z^{-1})^{i-1} \right] / D}{\left[ \sum_{i=1}^N \sqrt{D_i} (1 - z^{-1})^{i-1} \right] / D} \quad (13)$$

### NUMERICAL EXAMPLE

Results of computer simulation on a fourth order network function are presented.

#### Example 1.

The network function considered is

$$H(s) = \frac{[s^2 + 0.31(10^6)s + 0.003(10)^{12}]}{[s^4 + 0.804(10^6)s^3 + 1.4481(10^{12})s^2 + 0.009686(10^{18})s + 0.007056(10^{24})]} \quad (14)$$

$$= \frac{[s + 10^4][s + 0.3(10^6)]}{[s^2 + 0.004(10^6)s + 0.0049(10^{12})] [s^2 + 0.8(10)^6 s + 1.44(10)^{12}]}$$

s-poles:  $(-0.002 \pm j 0.0699714)(10^6)$   
 $(-0.400 \pm j 1.131371)(10^6)$

It was converted to a digital equivalent form (using pole-zero  $z = \text{Exp.}(sT)$  transform [11]) for computer simulation. With a sampling interval  $\Delta=0.5 \mu\text{s}$ . the z-domain transfer function turns out to be

$$H(z) = \frac{2.00z^{-2} - 3.7114409z^{-3} + 1.7128304z^{-4}}{1 - 3.379158z^{-1} + 4.428628z^{-2} - 2.718099z^{-3} + 0.6689807z^{-4}}$$

The system was excited was excited by a +-square 5  $\mu\text{s}$  pulse (see Fig. 2a). The model identified by the proposed method is

$$\hat{H}(z) = \frac{2.00z^{-2} - 3.71150z^{-3} + 1.7128z^{-4}}{1 - 3.3792z^{-1} + 4.4286z^{-2} - 2.7181z^{-3} + 0.66898z^{-4}}$$

s-poles:  $(-0.002 \pm j 0.0699714)(10^6)$   
 $(-0.399 \pm j 1.131373)(10^6)$

Using the inverse of the pole-zero transform, the s-domain transfer function can be obtained. The poles turn out as shown above.

The response of the model and the actual network response are compared in Fig. 2b.

### Example 2

The output of the network in Example 1 was corrupted by zero mean white noise such that the signal to noise ratio was 28 dB (see Fig. 3a). Ignoring the presence of noise the pencil of functions method was applied.

ignoring the presence of noise. The model identified is

$$\hat{H}(z) = \frac{2.0411z^{-2} - 3.7880z^{-3} + 1.7485z^{-4}}{1 - 3.60440z^{-1} + 5.1082z^{-2} - 3.4019z^{-3} + 0.89852z^{-4}}$$

s-poles:  $(-0.0011 \pm j 0.0702220)(10^6)$

$(-0.1059 \pm j 1.120712)(10^6)$

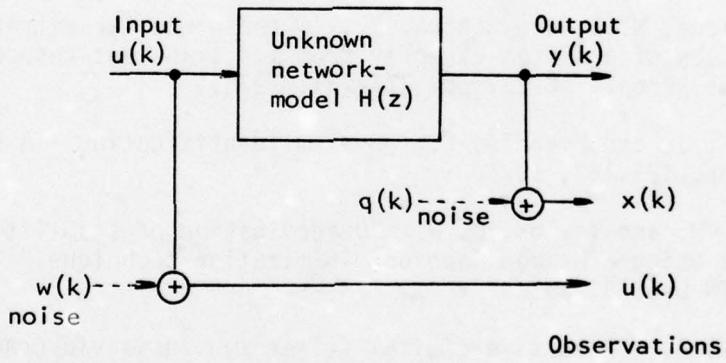
The responses of the model and the actual network response are compared in Fig. 3b. It is interesting to note that the imaginary parts of the poles have been identified quite accurately. However, the general degradation in the values of parameter estimates is obvious.

Two alternatives are available: a) we can accept the above model inasmuch as the above model behaves close to the original model, relative to the test input(s) used, or b) we may use noise corrections as described in [9] to obtain parameter estimates much closer to the original. The results of the second approach, although not presented here, do turn out close to the original model.

#### REFERENCES

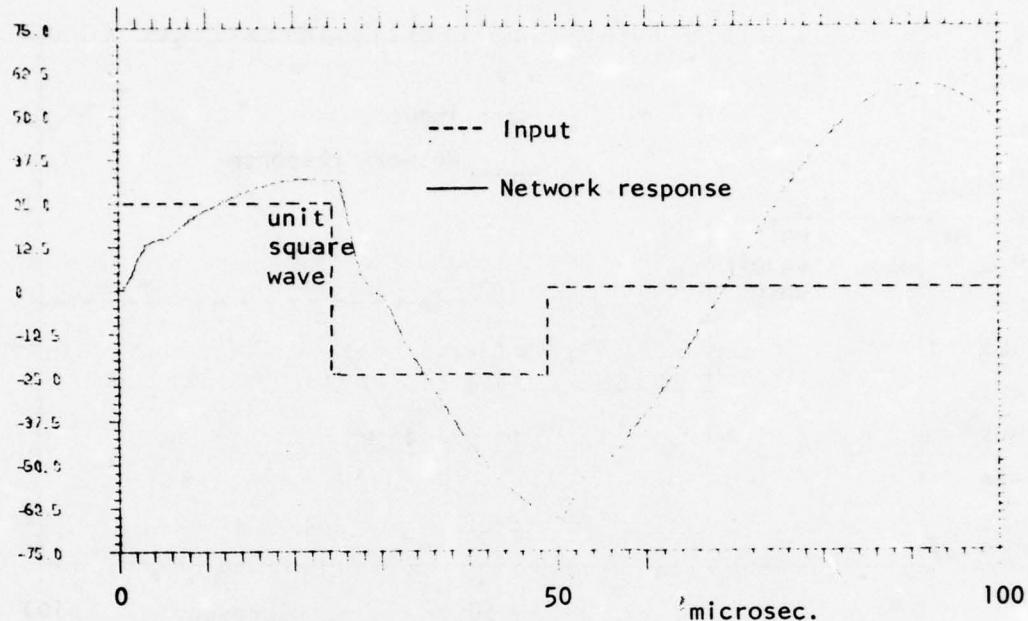
- [1] Van Blaricum, M. L. and Mittra, R., "A technique for extracting the poles and residues of a system directly from its transient response," IEEE Trans. Ant. Prop., AP-23, pp. 777-781, 1975.
- [2] Astrom, K. J. and Eykhoff, P., "System identification - A survey," Automatica, pp. 123-162, 1971.
- [3] Iliff, K. W. and Taylor, L. W., "Determination of stability derivatives from data using a Newton-Raphson minimization technique," NASA technical report, TN D-6579, 1972.
- [4] Cadzow, J. A., "Recursive digital filter synthesis via gradient based algorithms," IEEE Trans. Acous. Sp. Signal Proc., Vol. ASSP-24, pp. 349-355, October 1976.
- [5] Evans, E. G. and Fischl, "Optimal least-squares time-domain synthesis of recursive digital filters," IEEE Trans. Audio Electroacous., Vol. AU-21, pp. 61-65, February 1973.

- [6] Levin, M. J., "Estimation of a system pulse transfer function in the presence of noise," IEEE Trans. Autom. Control, Vol. AC-9, pp. 229-235, July 1964.
- [7] Kashyap, R. L., "Maximum likelihood identification of stochastic linear systems," IEEE Trans. Autom. Control, Vol. AC-15, pp. 25-34, February 1970.
- [8] Mehra, R. K., "Identification of stochastic linear dynamic systems using Kalman filter representation," AIAA J., Vol. 9, pp. 28-31, January 1971.
- [9] Jain, V. K., "Filter analysis by use of pencil-of-functions: Part I," IEEE Trans. Circuits and Sys., Vol. CAS-21, pp. 580-583, Sept. 1974.
- [10] Chenny, E. W., Introduction to Approximation Theory. New York: McGraw-Hill, 1966.
- [11] Stanley, W. D., Digital Signal Processing. Reston (Prentice-Hall): Reston, 1975.

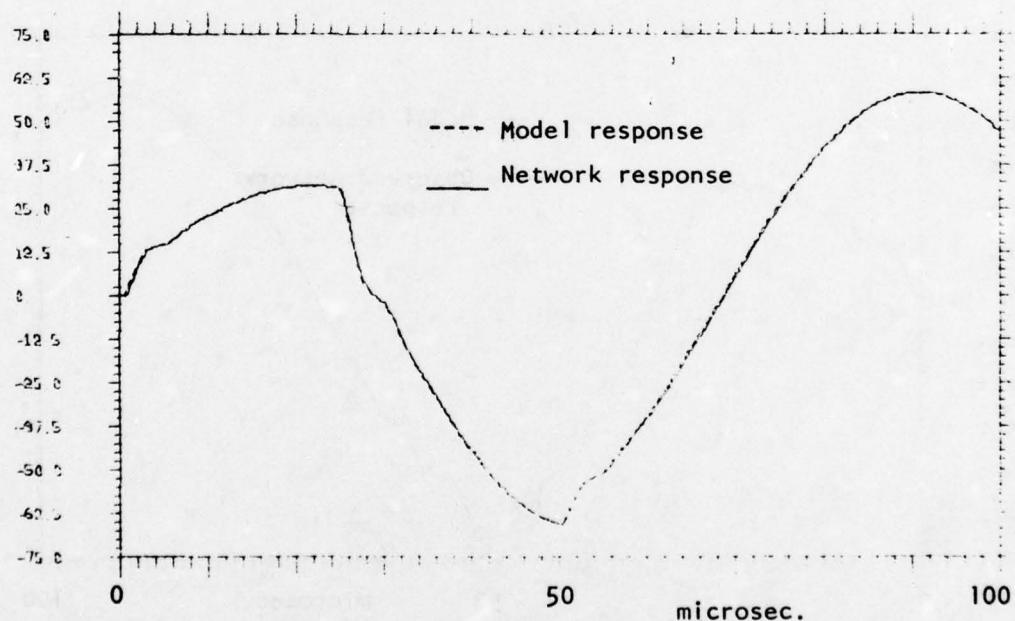


$$H(z) = \frac{Y(z)}{U(z)}$$

Fig. 1. Single input, single output identification problem.

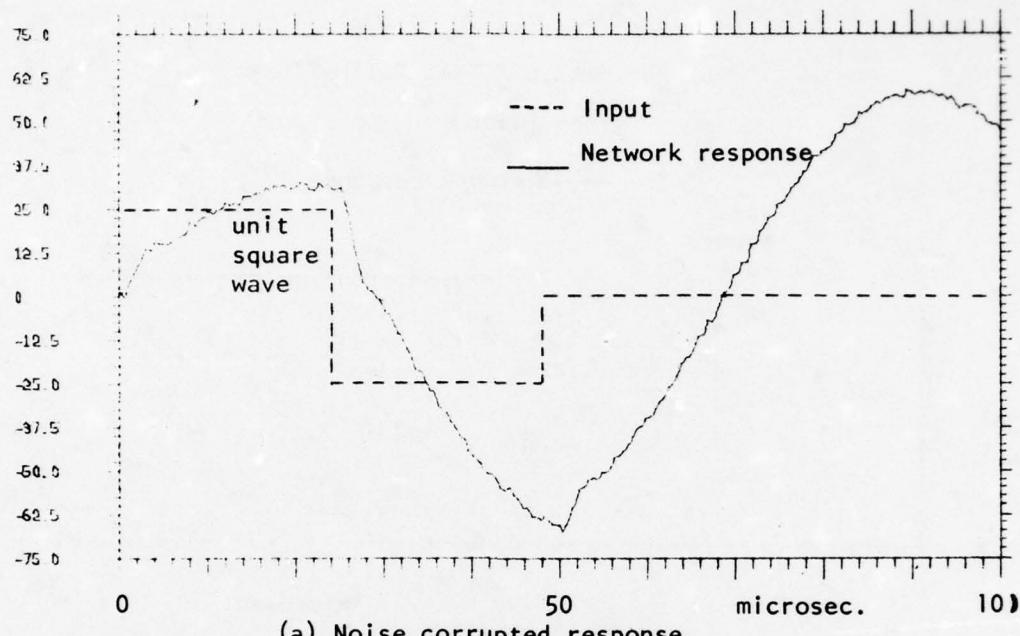


(a) input and response

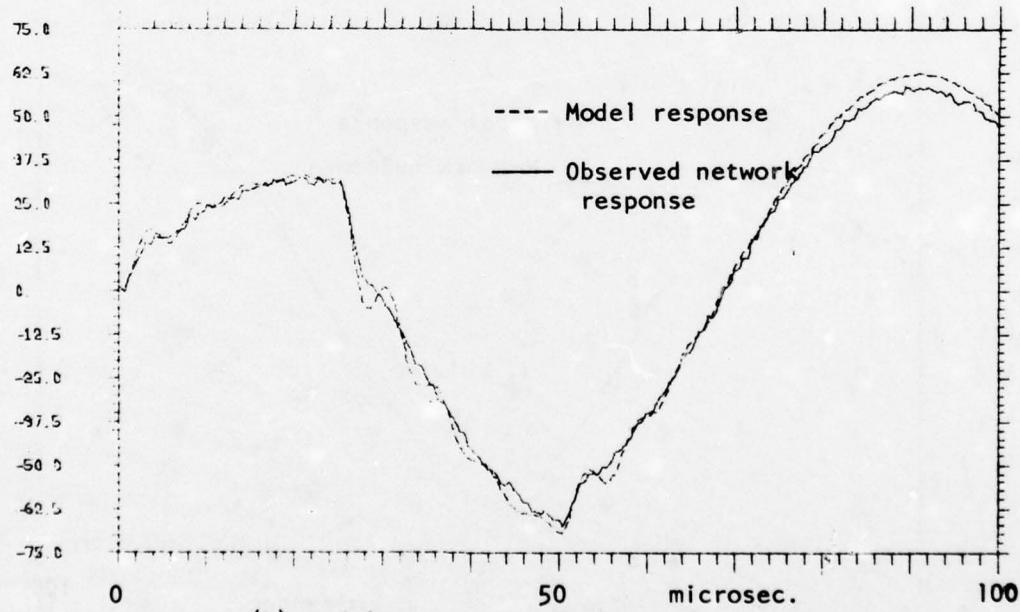


(b) model response and actual response

Fig. 2. Identification of a fourth order system  
(data uncorrected by noise)



(a) Noise corrupted response  
to square wave input



(b) Model response and observed response  
Fig. 3. Identification of a fourth order system  
(data corrupted by -28 dB noise)

## TWO DIMENSIONAL SPECTRAL ESTIMATION\*

ANIL K. JAIN  
SURENDRA RANGANATH

Department of Electrical Engineering  
State University of New York at Buffalo  
Bell Hall, Amherst Campus  
Buffalo, N.Y. 14260

### ABSTRACT

The problem of maximum entropy spectral estimation in two dimensions is considered. Unlike the one dimensional case, the two dimensional maximum entropy spectrum cannot be obtained by solving two dimensional linear equations of autoregression. An iterative algorithm is therefore considered. Examples on fields whose spectra is composed of discrete lines or is irrational are considered.

### INTRODUCTION

Two dimensional spectral estimation is the problem of estimating the spectral density function of a random field whose covariances (or autocorrelations) are given on a finite two dimensional window. Let  $\{u_{i,j}\}$  denote a zero mean, stationary, Gaussian random field whose covariance function is defined as

$$r(m,n) = E u_{i,j} u_{i+m,j+n} \quad (1)$$

Suppose  $r(m,n)$  is given on a window  $W = \{m,n; -p \leq m \leq p, -q \leq n \leq q\}$ . The spectral density  $S(f_1, f_2)$  of  $\{u_{i,j}\}$  is related to  $r(m,n)$  by

\*Presented at the Workshop on Spectral Estimation, held at Rome Air Development Center, Griffis Air Base, Rome, NY, May, 1978. Research supported in part by Rome Air Development Center, Contract F30602-75-C-0122 and Army Research Office, Research Triangle Park, North Carolina, under research grant No. DAAG29-77-G0044.

$$r(m,n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S(f_1, f_2) e^{j2\pi(mf_1 + nf_2)} df_1 df_2 \quad (2)$$

### MAXIMUM ENTROPY ESTIMATION

The maximum entropy estimate maximizes

$$H = \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln S(f_1, f_2) df_1 df_2 \quad (3)$$

under the linear constraints of (2) for  $m, n \in W$ . The result requires that the Fourier series of  $S^{-1}$  be truncated to terms within the window  $W$  i.e.,

$$S(f_1, f_2) = \beta^2 \left[ \sum_{m, n \in W} a_{m,n} e^{-j2\pi(mf_1 + nf_2)} \right]^{-1} \quad (4)$$

Thus, one has to solve for  $a_{m,n}$  given  $r_{m,n}$ . In one dimension this problem is equivalent to solving a set of linear equations of autoregression. However, this is not the case in two dimensions. Suppose we introduce a causal autoregressive representation

$$u_{i,j} = \sum_{m=0}^p \sum_{n=0}^q \alpha_{m,n} u_{i-m, j-n} + \epsilon_{i,j} \quad (5)$$

$(m,n) \neq (0,0)$

$$\text{where } E\epsilon_{i,j} = 0, \quad E\epsilon_{i,j}\epsilon_{i+m, j+n} = \beta^2 \delta_{m,0} \delta_{n,0} \quad (6)$$

i.e.,  $\{\epsilon_{i,j}\}$  is a white noise field. Then, given  $r_{m,n}$  for  $m, n \in W$ , one can find  $\alpha_{m,n}$ ,  $m, n \in W$ ,  $W = \{m, n; 0 \leq m \leq p, 0 \leq n \leq q\}$  by solving a set of linear equations

$$\mathcal{R}\alpha = -\beta^2 \mathbf{l}, \quad \beta^2 = 1 / [\mathcal{R}^{-1}]_{1,1} \quad (7)$$

where  $\mathcal{R}$  is a  $(p+1) \times (p+1)$  block Toeplitz covariance matrix of basic dimension  $(q+1)$ ,  $\alpha$  is a vector of unknowns  $\{\alpha_{m,n}\}$  with  $\alpha_{00} \triangleq -1$ , and  $\mathbf{l}$  is a unit

vector with 1 in the first location and zeros elsewhere. Solution of (7) for  $\alpha$  does not ensure a stable representation of (5) even if  $R$  is positive definite. Moreover, if the solution of (7) leads to a stable representation of (5), the covariances realized by (5) (for finite p,q) over the window W need not be the same as  $\{r(m,n)\}$  over W.

The spectral density realized by (5) is given by

$$S_u(f_1, f_2) = \beta^2 \left[ \sum_{m,n \in W} \hat{a}_{m,n} e^{-j2\pi(mf_1 + nf_2)} \right]^{-1} \quad (8)$$

$$\hat{a}_{m,n} = \hat{a}_{-m,-n}^\Delta = \sum_{i,j \in W} \alpha_{i,j} \alpha_{i+m, j+n} \quad (9)$$

where  $\alpha_{0,0} = -1$ , and  $\alpha_{i,j} = 0$  for  $i,j \notin W$ .

Although (4) and (8) are of the same form, i.e., (5) yields a maximum entropy spectrum, it does not necessarily give the maximum entropy spectrum of (4). In other words the coefficients  $\hat{a}_{m,n} \neq a_{m,n}$  because (8) would not satisfy (2) for  $m,n \in W$  even though  $\alpha$  is the solution of (7). Hence, the linear equation (7) does not yield the solution of (2) and (4)\*.

Note that in one dimension, the spectral estimation equation

$$r(m) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi mf} \left[ \sum_{n=-p}^p a_n e^{j2\pi nf} \right]^{-1} df \quad (10)$$

could be uniquely solved by solving for  $\alpha_m$  in the one dimensional analog of (5)

$$u_i = \sum_{m=1}^p \alpha_m u_{i-m} + \epsilon_i \quad (11)$$

via the linear equations

$$R\alpha = -\beta^2 \underline{1}, \quad \beta^2 = 1 / (R^{-1})_{1,1} \quad (12)$$

---

\*This statement contradicts the implications in [1] where the solution of (2) and (4) is proposed via (7) to (9).

where  $R$  is simply the  $(p+1) \times (p+1)$ , positive definite Toeplitz matrix,  $\{r(m-n)\}$ . Here, the positivity of  $R$  is enough to ensure stability of (11) and to yield from (11) the same covariances as the given  $\{r(m)\}$  for  $-p \leq m \leq p$ .

#### ITERATIVE SOLUTION

To obtain the spectrum  $S(f_1, f_2)$  we have to solve the nonlinear equations (2) and (4) for  $m, n \in W$ . To solve this problem we use an extension of an algorithm proposed by Censor et al; [2] addressed to maximizing  $\sum_j \ln x_j$  such that  $\bar{x} > 0$  and  $A\bar{x} = \bar{b}$  where  $A$  is an  $m \times n$  matrix. The solution  $\bar{x}$  is found to be

$$x_j = \frac{1}{\sum_i a_{i,j} u_i} \quad (13)$$

where  $\bar{u}$  is the dual vector of  $\bar{x}$ .

The iteration algorithm cyclically varies the co-ordinates of  $u$  so that only one component is changed and the dual functional decreases at each iteration step.

In our problem then, we have

$$S(k, l) = \frac{1}{\sum_m \sum_n a_{m,n} w_N^{mk} w_N^{nl}} ; \quad w_N = e^{-j\frac{2\pi}{N}} \quad (14)$$

where we could consider  $a_{m,n}$  as our dual variables.

In the algorithm we cyclically vary  $m$  and  $n$  over the window  $W$  so that only one component  $a_{i,j}$  is changed at each iteration.

At the  $(q+1)^{th}$  iteration, where we change  $a_{i,j}$

$$\begin{aligned} S_{k,l}^{q+1} &= \frac{1}{\sum_m \sum_n a_{m,n}^{q+1} w_N^{(mk + nl)}} \\ &= \frac{1}{\sum_m \sum_n a_{m,n}^q w_N^{(mk + nl)} + a_{ij}^{q+1} w_N^{(ik + jl)}} \\ &\quad (m,n) \neq (i,j) \end{aligned}$$

$$= \frac{1}{\sum_{m,n \in W} \sum_{m,n} a_{m,n}^q w_N^{mk} + nl + w_N^{(ik+jl)} [a_{ij}^{q+1} - a_{ij}^q]} \quad (15)$$

Hence,

$$s_{k,l}^{q+1} = \frac{s_{k,l}^q}{1 + s_{k,l}^q [a_{ij}^{q+1} - a_{ij}^q]} \cdot \quad (16)$$

The value of  $a_{ij}^{q+1} - a_{ij}^q$  is calculated from the constraint

$$f(z) = r(i,j) - \frac{1}{2} \sum_{N} \sum_{k,l} s_{k,l}^{q+1} w_N^{-ik} w_N^{-jl} = 0 \quad (17)$$

This equation can be solved for  $z = a_{ij}^{q+1} - a_{ij}^q$  in a few iterations of the Newton-Raphson technique. The function  $f(z)$  has many zeros and it must be insured that we pick that value of  $z$  that satisfies

$$1 + s_{k,l}^q w_N^{-ik} w_N^{-jl} z > 0$$

The equations involved are all reduced to real equations by symmetry properties i.e.,

$$a_{mn} = a_{-m-n}; a_{m,-n} = a_{-m,n}$$

and

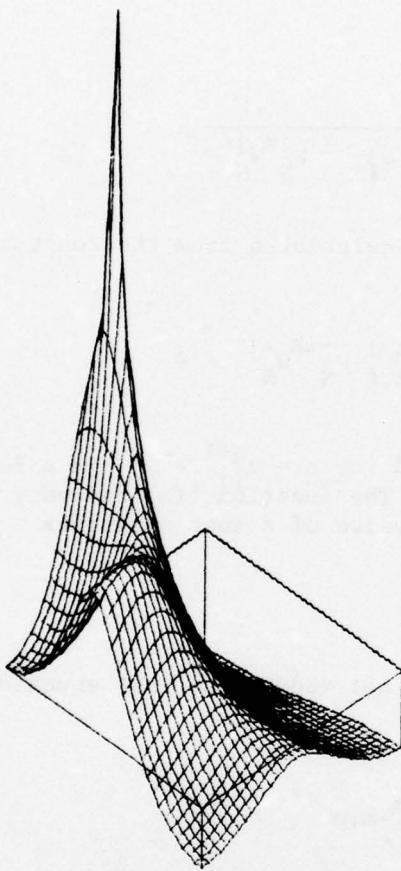
$$s_{k,l} = s_{-k,-l} \text{ and } s_{-k,l} = s_{k,-l} .$$

#### EXAMPLE

We considered the following covariance model [3]

$$r(k,l) = \cos 2\pi(0.05k + 0.2l) + 0.5 \cos 2\pi(0.2k + 0.05l) + 0.2\delta(k,l)$$

which yields a spectra of two lines (delta functions) and a constant in the 2-dimensional frequency plane.



ESTIMATED SPECTRUM OF TWO  
COSINE WAVES IN 10% NOISE

FIGURE 1

Figure 1 shows the estimated spectra where the larger peak has an amplitude of 35dB and the smaller an amplitude of 8 dB. Other examples and numerical tradeoffs of the proposed algorithm will be presented at the conference.

#### REFERENCES

1. W.I. Newman, "A New Method of Multidimensional Power Spectral Analysis", Astrm. Astrophys 54, pp. 369-380, 1977.
2. Y. Censor, etal; "Relaxational Methods for Large Scale Entropy Optimization Problems, with Application in Image Reconstruction", presented at the Proceedings of The Information Linkage between Applied Mathematics and Industry Workshop, Feb. 1978, Monterey, California.
3. J.W. Woods, "Two Dimensional Markov Spectral Estimation", IEEE Transactions on Information Theory, Vol. IT-22, #5, September 1976.

# LATTICE METHODS IN SPECTRAL ESTIMATION

JOHN MAKHOUL

Bolt Beranek and Newman Inc.  
Cambridge, MA 02138

## Abstract

The lattice structure offers a convenient visual realization of the Levinson recursion. Based on the lattice, a number of methods have been developed recently for all-pole or autoregressive spectral estimation. Of particular importance are methods that do not require windowing of the data, especially for short data lengths. This paper focuses on the importance of the lattice as a tool in spectral estimation. Various lattice methods are presented and compared, including adaptive estimation.

## 1. Introduction

All-pole or autoregressive spectral estimation, which is a special case of linear prediction analysis, has become popular as a simple, effective method of spectral estimation. In this paper, we first review the two basic methods of linear prediction analysis, the autocorrelation and covariance methods [1], where the first assumes that the signal is windowed (i.e., the signal is zero outside the range of interest), and the second makes no assumptions about the signal outside the range of interest. Then, we present the lattice structure and the various methods that can be derived from it. We shall see that the methods of Itakura [2] and Burg [3] are special cases of the general class of methods to be presented. Cases where the latter two methods give incorrect results will be given, as well as suggestions of which methods might be more appropriate.

In the sequel we shall assume that the signal spectrum is to be modelled by an all-pole transfer function

$$H(z) = G/A(z) \quad (1)$$

where  $A(z) = 1 + \sum_{k=1}^p a(k)z^{-k}$  (2)

is the "inverse filter" and  $G$  is a gain factor. Whenever the word "stable" is used below, it shall refer to the stability of the filter  $H(z)$ , implying that  $H(z)$  and  $A(z)$  are minimum-phase, and that the poles of  $H(z)$  (and zeros of  $A(z)$ ) are inside the unit circle.

## 2. Linear Prediction

In all-pole linear prediction analysis we assume that the signal  $x(n), 0 \leq n \leq N-1$ , can be approximated as a weighted linear summation of past samples:

$$x(n) \hat{x}(n) = - \sum_{k=1}^p a(k)x(n-k) \quad (3)$$

where  $a(k)$  are the predictor coefficients and  $p$  is the order of the predictor. The coefficients  $a(k)$  are computed as the result of minimizing the total squared value of the error  $e(n) = x(n) - \hat{x}(n)$ :

$$E = \sum_n e^2(n) = \sum_n [x(n) + \sum_{k=1}^p a(k)x(n-k)]^2. \quad (4)$$

There are two cases of interest.

### a) Windowed Case

Here we assume that  $x(n)=0$  for  $n < 0$  and  $n > N$ ; then minimization of (4) results in the normal equations of the "autocorrelation method":

$$\sum_{k=1}^p a(k)R(i-k) = -R(i), \quad 1 \leq i \leq p, \quad (5)$$

and  $E_p^* = R(0) + \sum_{k=1}^p a(k)R(k) \quad (6)$

where  $R(i) = \sum_{n=|i|}^{N-1} x(n)x(n-|i|) \quad (7)$

is the signal autocorrelation, and  $E_p^*$  is the minimum total error (or residual). (The shape of the window to be used in multiplying the signal is of importance, but will not be discussed here.)

The autocorrelation matrix  $\{R(i-k)\}$  in (5) is Toeplitz, and the set of equations may be solved using the algorithm developed by Levinson, Robinson and Durbin [1]. We shall simply refer to it as the Levinson recursion:

$$E_0 = R(0) \quad (8a)$$

$$K_m = -[R(m) + \sum_{k=1}^{m-1} a_{m-1}(k)R(m-k)]/E_{m-1}^* \quad (8b)$$

$$a_m(m) = K_m \quad \left. \begin{array}{l} \\ a_m(k) = a_{m-1}(k) + K_m a_{m-1}(m-k), \quad 1 \leq k \leq m-1 \end{array} \right\} \quad (8c)$$

$$E_m^* = (1 - K_m^2) E_{m-1}^* \quad (8d)$$

Equations (8b)–(8d) are solved recursively for  $m=1, 2, \dots, p$ . The final solution is given by  $a(k)=a_p(k), 1 \leq k \leq p$ . Note that in obtaining the solution for order  $p$ , one computes the solutions for all predictors of order less than  $p$ . We shall see below how (8) is implemented in lattice form. In particular, the reflection or partial-correlation coefficients  $K_m$  will be the parameters of the lattice. One can show that the following condition for  $K_m$  in (8b) holds:

$$|K_m| < 1, \quad 1 \leq m \leq p, \quad (9)$$

which guarantees the stability of the filter  $H(z)$ .

### b) Unwindowed Case

Here we make no assumptions about the signal outside the given range. From (4) one can see that in order not to need data outside the given interval, the summation should go from  $n=p$  to  $n=N-1$ . The minimization of (4) results in the normal equations of the "covariance method":

$$\sum_{k=1}^p a(k) R(k,i) = -R(\emptyset, i), \quad 1 \leq i \leq p, \quad (10)$$

and  $E_p^* = R(\emptyset, \emptyset) + \sum_{k=1}^p a(k) R(\emptyset, k), \quad (11)$

where  $R(k,i) = \sum_{n=p}^{N-1} x(n-k)x(n-i) \quad (12)$

is the covariance of the signal. The covariance matrix  $\{R(k,i)\}$  is not Toeplitz and so (10) cannot be solved using the Levinson recursion. Note that the covariance method is reminiscent of Prony's method.

### Comparison

In the autocorrelation (windowed) method, the resulting all-pole filter is stable, at a cost of windowing of the signal, which results in a loss of frequency resolution, especially for small  $N$ . On the other hand, the covariance (unwindowed) method has the advantage of no windowing, but filter stability is not guaranteed, especially for small  $N$ . We now present the lattice methods and see how some of these problems may be alleviated.

### 3. Lattice Formulation

The lattice was developed by Itakura [2] for use in the analysis and synthesis of speech. The development here follows that of the author [4]. Fig. 1 shows the basic lattice used in the analysis. From Fig. 1, the following relations hold:

$$f_\emptyset(n) = g_\emptyset(n) = x(n) \quad (13a)$$

$$f_m(n) = f_{m-1}(n) + K_m g_{m-1}(n-1) \quad (13b)$$

$$g_m(n) = K_m f_{m-1}(n) + g_{m-1}(n-1) \quad (13c)$$

where  $f_m(n)$  is the "forward" residual at stage  $m$ , and  $g_m(n)$  is the "backward" residual. Taking the z-transform of (13), one obtains a similar recursion with  $F_m(z)$  and  $G_m(z)$  as the z-transforms of  $f_m(n)$  and  $g_m(n)$ , respectively. Defining the transfer functions to stage  $m$  as

$$A_m(z) = F_m(z)/X(z), \quad B_m(z) = G_m(z)/X(z) \quad (14)$$

one obtains the following recursive relations:

$$A_0(z) = B_0(z) = 1 \quad (15a)$$

$$A_m(z) = A_{m-1}(z) + K_m z^{-1} B_{m-1}(z) \quad (15b)$$

$$B_m(z) = K_m A_{m-1}(z) + z^{-1} B_{m-1}(z). \quad (15c)$$

Let the forward transfer function at stage  $m$  be given by

$$A_m(z) = \sum_{k=0}^m a_m(k) z^{-k}, \quad (16a)$$

then from (15) one can show that  $B_m(z)$  will be the corresponding reverse polynomial:

$$B_m(z) = z^{-m} A_m(z^{-1}) = \sum_{k=0}^m a_m(m-k) z^{-k}. \quad (16b)$$

From (15) and (16) we also have:

$$a_m(0) = 1, \quad a_m(m) = K_m. \quad (17)$$

By comparing (15) and (17) with (8c), we see that (15) is the heart of the Levinson recursion.

Given the polynomial  $A_p(z)$ , with  $a_p(0)=1$ , one can generate all the polynomials  $A_m(z)$ ,  $m < p$ , and the coefficients  $K_m$ , using the following reverse recursion derived from (15):

$$A_{m-1}(z) = [A_m(z) - K_m B_m(z)] / (1 - K_m^2) \quad (18)$$

along with (16b).  $A_p(z)$  and  $K_m$ ,  $1 \leq m \leq p$ , are uniquely related iff  $A_p(z)$  is minimum phase and, hence, (9) is true. In that case,  $B_p(z)$  is maximum phase, as one can see from (16b).

The all-pole modelling problem can now be stated as the minimization of the output residual energy with respect to the reflection coefficients  $K_m$ . One can then use the recursion (15) to determine the predictor coefficients. Again, here, we have two cases: windowed and unwindowed.

#### 4. Windowed Case

For the case where the data is stationary or windowed, one can show that, at each stage, the energy of the forward residual is equal to the backward residual energy:

$$\overline{E_m} = \overline{f_m^2(n)} = \overline{g_m^2(n)}, \text{ for all } m, \quad (19)$$

where the overbar denotes summation over  $n$ , in this case over all time. From (13b) and (19), one can show that

$$E_m = E_{m-1} + 2K_m \overline{f_{m-1}(n) g_{m-1}(n-1)} + K_m^2 E_{m-1}. \quad (20)$$

By defining

$$r_{m-1} = \frac{\overline{f_{m-1}(n) g_{m-1}(n-1)}}{\overline{E_{m-1}}} \quad (21)$$

we have

$$E_m = (1 + 2r_{m-1}K_m + K_m^2) E_{m-1}. \quad (22)$$

From (19), one can rewrite (21) as

$$r_{m-1} = \frac{\overline{f_{m-1}(n) g_{m-1}(n-1)}}{\sqrt{\overline{f_{m-1}^2(n)} \overline{g_{m-1}^2(n-1)}}} \quad (23)$$

which is the correlation coefficient between  $f_{m-1}(n)$  and  $g_{m-1}(n-1)$ , the inputs to stage  $m$ . Therefore, (22) gives the energy at stage  $m$  as a function of the energy at stage  $m-1$ , the correlation coefficient  $r_{m-1}$ , and  $K_m$ .

For  $p > 1$ , the residuals are fairly complicated functions of  $K_m, 1 \leq m \leq p$ , and the minimization of  $E_p$  leads to a nonlinear minimization problem. However, for a windowed or stationary signal, the  $p$ th-order global nonlinear minimization problem can be solved as a sequence of 1st-order local minimization problems at each stage. We shall call this major property as the lattice minimization property, or simply LM property. In the implementation, if we assume that  $E_{m-1}$  is minimized as  $E_{m-1}^*$ , then the minimization of  $E_m$  in (22) is accomplished by differentiating with respect to  $K_m$  to obtain the optimal value

$$K_m^* = -r_{m-1}, \quad (24)$$

and  $E_m^* = (1 - K_m^{*2}) E_{m-1}^*. \quad (25)$

Equation (24) is the reason why  $\{K_m\}$  are known as partial correlation coefficients. Since  $|r_{m-1}| \leq 1$ , so is  $K_m^*$ , and  $E_m^* \leq E_{m-1}^*$ . Equation (25) is identical to (8d), and (24) with (21) can be shown to be identical to (8b).

As a result of the minimization, one can show that the backward residuals become orthogonal to each other:

$$\overline{g_i(n) g_j(n)} = E_i^* \delta_{ij}. \quad (26)$$

Other correlation properties are given in [4].

Equation (26) is a very important property, which has been very useful in adaptive lattice estimation (see Section 7).

### 5. Unwindowed Case - Suboptimal Solutions

We have seen, in the windowed case, that the LM property allowed us a recursive computation of  $K_m$ . Unfortunately, the LM property is not true if the signal is not windowed. The nonlinear minimization problem cannot be replaced by a sequence of one-stage minimizations. However, one can obtain a suboptimal solution by pretending that the LM property does in fact hold. Due to the suboptimality of such a solution, one can define a large number of suboptimal solutions, as we shall see below.

Two reasonable methods are to minimize either the forward or backward residual energy at each stage. Define

$$\bar{F}_m = \overline{f_m^2(n)} \quad (27a)$$

$$\bar{G}_m = \overline{g_m^2(n)} \quad (27b)$$

where the overbar now denotes summation over a fixed finite interval. By substituting (13b) in (27a) and (13c) in (27b) and differentiating with respect to  $K_m$ , we obtain two different solutions:

$$K_m^f = - \frac{\overline{f_{m-1}(n) g_{m-1}(n-1)}}{\overline{g_{m-1}^2(n-1)}} \quad (28a)$$

$$K_m^g = - \frac{\overline{f_{m-1}(n) g_{m-1}(n-1)}}{\overline{f_{m-1}^2(n)}} \quad (28b)$$

It is clear that  $K^f$  and  $K^g$  have the same sign  $S$ :

$$S = \text{sign } K^f = \text{sign } K^g. \quad (29)$$

One important fact is that  $K^f$  and  $K^g$  as defined in (28) need not obey (9). However, one can also show that if the magnitude of either of them is greater than 1, the magnitude of the other is necessarily less than 1 [5]:

If  $|K^f| > 1$ , then  $|K^g| < 1$   
 or if  $|K^g| > 1$ , then  $|K^f| < 1$ . (30)

Now, define the generalized  $r$ th mean between  $K^f$  and  $K^g$ :

$$K^r = S \left[ \frac{1}{2} (|K^f|^r + |K^g|^r) \right]^{1/r}. \quad (31)$$

One can show that [5]

$$|K^r| < 1, \text{ iff } r \leq 0. \quad (32)$$

In particular, as  $r \rightarrow 0$ , we obtain the geometric mean

$$K_m^0 = K_m^I = - \frac{\bar{f}_{m-1}(n) g_{m-1}(n-1)}{\sqrt{\bar{f}_{m-1}^2(n) \bar{g}_{m-1}^2(n-1)}} \quad (33)$$

which is the negative of the correlation between  $f_{m-1}(n)$  and  $g_{m-1}(n-1)$ . This is the equation used by Itakura [2]. For  $r = -1$ , we have the solution proposed by Burg [3]:

$$K_m^{-1} = K_m^B = - \frac{2 \bar{f}_{m-1}(n) g_{m-1}(n-1)}{\bar{f}_{m-1}^2(n) + \bar{g}_{m-1}^2(n)} \quad (34)$$

Equation (34) may also be obtained by minimizing the sum of the forward and backward residual energies. Another solution, proposed by the author, is obtained as  $r \rightarrow -\infty$ , which gives the minimum magnitude:

$$K^{-\infty} = K^M = S \min [|K^f|, |K^g|]. \quad (35)$$

One can show that

$$|K^M| \leq |K^B| \leq |K^I|. \quad (36)$$

Therefore, there is a host of suboptimal solutions to choose from, all of which obey (9). Whichever method one chooses, the procedure for the solution is as follows: Compute  $K_1$ , then evaluate  $f_1(n)$  and  $g_1(n)$  in the range of interest; now compute  $K_2$ , then evaluate  $f_2(n)$  and  $g_2(n)$ , etc. While this procedure is straightforward, it can be expensive computationally, especially for large  $N$ . We now present an alternate implementation that is more efficient.

#### Efficient Implementation

We first compute the signal covariances  $R(k, i)$  from (12). The remainder of the solution uses  $R(k, i)$  and  $a_m(k)$  for  $m=1, 2, \dots, p$ . From (14) and (16), one can write

$$f_m(n) = \sum_{k=0}^m a_m(k)x(n-k) \quad (37a)$$

$$g_m(n) = \sum_{k=0}^m a_m(m-k)x(n-k). \quad (37b)$$

One can easily show that [5]

$$\overline{f_{m-1}(n)g_{m-1}(n-1)} = \sum_{k=0}^m \sum_{i=0}^m a_{m-1}(k)a_{m-1}(i)R(k,m-i) \quad (38a)$$

$$\overline{f_{m-1}^2(n)} = \sum_{k=0}^m \sum_{i=0}^m a_{m-1}(k)a_{m-1}(i)R(k,i) \quad (38b)$$

$$\overline{g_{m-1}^2(n-1)} = \sum_{k=0}^m \sum_{i=0}^m a_{m-1}(k)a_{m-1}(i)R(m-k,m-i). \quad (38c)$$

These values can be used with any of the solutions for  $K_m$ . Therefore, the procedure now is as follows: Compute  $K_1$ , then  $A_1(z)$  using (8c) or (15b); evaluate the terms in (38), compute  $K_2$ , then  $A_2(z)$ , etc. The computations in (38) can also be easily cut by about one-half [5]. The procedure given above results in about a four-fold computational savings over the methods of Itakura and Burg.

### Comparison

If the signal is windowed, then all the solutions presented above give the same optimal solution. For an unwindowed signal, the various definitions of  $K_m$  give different suboptimal solutions. If the signal is a sample of a random process or a periodic signal, and  $N$  is large, the various definitions give similar results. The problem arises for small  $N$ , or for certain deterministic signals, as we shall see in the examples below.

Example 1 - Let  $x(n)$  be the impulse response of a single-pole filter  $1/(1-cz^{-1})$ , where  $0 < c < 1$ , i.e.  $x(n)=c^n$ ,  $n \geq 0$ , and  $x(n)=0$ ,  $n < 0$ . Assume that we are given three samples ( $N=3$ ):  $x(0)=1$ ,  $x(1)=c$ , and  $x(2)=c^2$ . We wish to compute the single parameter of  $A_1(z)=1+a_1(1)z^{-1}$ , i.e.  $p=1$ . From (28) and (38), we have:

$$K_1^F = -R(0,1)/R(1,1); \quad K_1^G = -R(0,1)/R(0,0). \quad (39)$$

Using (12), we have for the different solutions for  $K_1$ :

$$K_1^F = -c; \quad K_1^G = -1/c \quad (40a)$$

$$K_1^I = 1; \quad K_1^B = -\frac{2c}{1+c}; \quad K^M = -c. \quad (40b)$$

Note that, of the three solutions in (40b), only  $K^M$  gives the correct result. As  $c \rightarrow 1$ , all solutions approach 1. This example shows that the popular use of  $K^B$  may not give the correct result. One may conclude then, that minimizing the sum of the forward and backward residual energies may not always be advisable.

Example 2. Let  $x(n)$  be the impulse response of a 2-pole filter  $\frac{1}{1-z^{-1}+z^{-2}}$ . For this case,  $x(n)$  is a sine wave with a frequency of  $1/6$  Hz (sampling rate = 1 Hz). One can show that  $x(n) = \sin[(n+1)\pi/3]/\sin(\pi/3)$ , for  $n > 0$ . The first few samples starting with  $n=0$  are:  $1, 1, 0, -1, -1, 0, \dots$ . For  $p=2$ ,  $A_2(z)$  is of the form:

$$A_2(z) = 1 + K_1(1+K_2)z^{-1} + K_2z^{-2}. \quad (41)$$

Therefore, the desired values of the predictor coefficients are  $a_2(1)=-1$  and  $a_2(2)=1$ , from which we find that  $K_1=-1/2$  and  $K_2=1$ . Assume now that we are given only the first four points ( $N=4$ ), which would normally be sufficient to determine the correct coefficient values. For  $N=4$ , we find using (12) that  $R(0,1)=0$ , and therefore all the definitions of  $K_1$  will be equal to zero, instead of  $-1/2$ . Using (38) and (12), one can show that

$$K_2^F = 1/2 ; K_2^Q = 1 \quad (42a)$$

$$K_2^I = 1/\sqrt{2} ; K_2^B = 2/3 ; K_2^M = 1/2. \quad (42b)$$

Note that  $K_2^Q$  gives the only correct value, with  $K_2^I$  being the closest of the three solutions in (42b). However, since  $K_1=0$ , the whole answer is not even close to being correct, because even if we set  $K_2=1$ , we will have  $A_2(z)=1+z^{-2}$ , which is a sine wave of frequency  $1/4$  Hz. For  $K_2<1$ , the sine wave will decay, i.e. the poles will have a finite bandwidth.

This example has demonstrated that, irrespective of which definition of  $K_m$  one adopts, the stage-by-stage minimization may produce drastically erroneous results. The author believes that it is this problem that lies behind the "line shifting" and "line splitting" phenomenon observed when using Burg's method [6].

## 6. Unwindowed Case - Optimal Solution

A cure for the problems just presented is to perform the optimal minimization of the forward residual. The covariance method presented in Section 2, equations (10)-(12), give the optimal result. If the resulting all-pole filter is stable, then the answer will be identical to that obtained by the nonlinear minimization in the lattice. One can easily test the stability of such a solution by using the reverse recursion (18) and checking to see if all  $K_m$  obey (9). If so, then there is no need to use the lattice. If, on the other hand, the resulting filter is unstable, then we need to use the lattice. In this case, we perform a constrained minimization using (9) as our constraint. The problem is greatly simplified by transforming the constrained problem into an unconstrained problem. The idea is to perform an appropriate transformation on  $K_m$ ,  $L_m=T(K_m)$ , and minimize with respect to  $L_m$  instead [6]. Assuming we wish to minimize  $F_p$  from (27a), the gradient is then given by

$$\frac{\partial F_p}{\partial L_m} = \frac{\partial F_p}{\partial K_m} \frac{dK_m}{dL_m} . \quad (43)$$

Two transformations are of particular interest:

$$L' = \sin^{-1} K ; \frac{dK}{dL} = \sqrt{1 - K^2} \quad (44a)$$

$$L'' = \frac{1}{2} \log \frac{1+K}{1-K} = \tanh^{-1} K ; \frac{dK}{dL} = 1 - K^2 \quad (44B)$$

Note that whatever the values of  $L'$  and  $L''$  are,  $K$  is automatically constrained to obey (9).  $L'$  is a multi-valued function of  $K$ , while  $L''$  is a single-valued function of  $K$ , which might be preferable. As  $K \rightarrow 1$ , both transformations will probably lead to problems. Fougere [6] has circumvented these problems when using  $L'$  by setting  $K = U \sin L'$ , where  $U$  is a number very close but less than 1. One, therefore, might also use  $K = U \tanh L''$  in the same way. To the author's knowledge, no one has used  $L''$  in this application yet, though it has been useful for quantization purposes [7].

In a gradient search algorithm, the updating equation is usually of the form

$$L_m(i+1) = L_m(i) - \alpha_m \left. \frac{\partial F_p}{\partial L_m} \right|_{L_m=L_m(i)} \quad (45)$$

where  $i$  is the iteration number. Therefore, we need to compute (43), and hence need

$$\frac{\partial F_p}{\partial K_m} = 2 f_p(n) \frac{\partial f_p(n)}{\partial K_m} . \quad (46)$$

The lattice is very useful in evaluating (46). One can show that (46) is easily implemented as shown in Fig. 2. Again, great savings in computation can be obtained by making use of the covariance of the input, as we did in (38).

Fougere [6] has had good results minimizing the sum of the forward and backward residual energies. However, his problem was that of a sine wave, where the poles are close to the unit circle, in which case any type of lattice optimal minimization would do. In general, our experience tells us that minimizing only the forward residual energy is preferable. This also cuts the computational cost.

## 7. Adaptive Estimation

For applications where the short-term spectrum changes as a function of time (such as speech and line-tracking), the lattice offers a simple, fast-converging adaptive structure that has given results superior to the traditional adaptive transversal

filter. In particular, the adaptive lattice seems to converge with a speed that is independent of the eigenvalue spread of the input signal [9]. This can be shown to be due to the orthogonality property (26) of the backward residuals [4,8].

In a time-varying situation, one is mainly interested in the most recent history of the signal. Therefore, it is reasonable to weight the residuals such that the more recent values are given more importance. This can be accomplished by using a weighting window, to be distinguished from the data window mentioned in Section 2. Below, we shall use the definition in (34) for  $K_m$ ; similar equations can be written using other definitions.

Given  $K_m(n)$ ,  $1 \leq m \leq p$ , at time  $n$ , and the forward and backward residuals up to time  $n$ , the reflection coefficients at time  $n+1$  are computed from [8]

$$K_m^B(n+1) = - \frac{\sum_{k=-\infty}^n w(n-k) f_{m-1}(k) g_{m-1}(k-1)}{\sum_{k=-\infty}^n w(n-k) [f_{m-1}^2(k) + g_{m-1}^2(k-1)]} \quad (47a)$$

$$= - \frac{C_m(n)}{D_m(n)} \quad (47b)$$

where  $w(n)$  is the weighting window. The value of  $K_m$  in (47) is always guaranteed to obey (9). Equation (47) is computed for  $1 \leq m \leq p$ , then the new residuals at time  $n+1$  are computed, etc.

#### Weighting Windows

Data windows may be quite arbitrary, and may take on positive and negative values. In contrast, weighting windows must always be nonnegative [8]. In particular, we must have

$$\begin{aligned} w(n) &> 0, n > 0 \\ w(n) &= 0, n \leq 0. \end{aligned} \quad (48)$$

As examples, we present two types of windows: nonrecursive and recursive. The nonrecursive window is the usual rectangular window of width  $M$ :

$$\begin{aligned} w_1(n) &= 1, 0 \leq n \leq M-1, \\ &= 0, \text{ otherwise.} \end{aligned} \quad (49)$$

This window has some bad effects as a data window but has good properties as a weighting window. The recursive window is the impulse response of a single real pole:

$$w_2(n) = \begin{cases} \beta^n, & n \geq 0, \quad 0 < \beta \leq 1 \\ 0, & n < 0. \end{cases} \quad (50)$$

For this window, as well as other recursive windows, one can compute  $C_m$  and  $D_m$  in (47b) recursively. For  $w_2(n)$  we have from (47) and (50):

$$C_m(k) = \beta C_m(k-1) + 2f_{m-1}(k)g_{m-1}(k-1) \quad (51a)$$

$$D_m(k) = \beta D_m(k-1) + f_{m-1}^2(k) + g_{m-1}^2(k-1). \quad (51b)$$

Other all-pole windows may be defined, but because of condition (48), all such weighting windows must be the impulse responses of filters with positive real poles.

For the special window  $w_2(n)$ , one can show that (47a) can be written as:

$$K_m^B(n+1) = K_m^B(n) - \frac{f_{m-1}(n)g_m(n) + g_{m-1}(n-1)f_m(n)}{D_m(n)} \quad (52)$$

This is similar to the steepest gradient algorithm of Griffiths [9]. Because of the equality of (52) to (47a),  $K_m$  in (52) is guaranteed to obey (9) always. We point out that, with recursive windows, one might be able to use  $K^f$  instead of  $K^B$  without fear that (9) would be violated.

#### Spectral Estimation Example

The data  $x(n)$  was generated by passing white Gaussian noise through an 11-pole filter whose transfer function is shown in Fig. 3. The adaptive estimation procedure shown in (47b) and (51) with  $\beta=1$  was used in order to test the convergence speed of the lattice. The goodness of fit after each iteration  $j$  was measured using the likelihood test of Itakura [10]:

$$I(j) = \frac{1}{E_p^*} \sum_{k=0}^p \sum_{i=0}^p a_p^j(k) a_p^j(i) R(i-k) \quad (53)$$

where  $E_p^*$  is the minimum possible residual energy as computed from (25),  $R(i)$  is the autocorrelation corresponding to the power spectrum in Fig. 3, and  $a_p^j(k)$  are the predictor coefficients computed from the reflection coefficients at iteration  $j$ .  $I(j)$  is always greater or equal to 1; it is equal to 1 iff the computed filter is identical to the desired filter. Fig. 4a shows a plot of  $I(j)$  as a function of  $j$  for a single record. Fig. 4b shows the average of estimates corresponding to 15 different records. Note the speed of convergence; it is generally proportional to  $p$ , but is not affected by the spectral dynamic range of  $x(n)$  (i.e., is not affected by the eigenvalue spread, which is about 1000 in this example).

## 8. Conclusion

The purpose of this paper was to show that the lattice is a very useful tool in spectral estimation, especially for short data lengths. The orthogonality property of the backward residuals has given the lattice fast convergence properties that are especially useful also in adaptive Wiener filtering, such as in adaptive equalizers [4,8], and adaptive noise cancelling [9].

## Acknowledgment

The author wishes to thank his colleague R. Viswanathan for discussions on adaptive lattice estimation and for providing him with the results in Figs. 3 and 4 of this paper.

## References

1. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, pp. 561-580, April 1975.
2. F. Itakura and S. Saito, "Digital Filtering Techniques for Speech Analysis and Synthesis," Proc. 7th Int. Cong. Acoust., Budapest, Paper 25-C-1, pp. 261-264, 1971.
3. J.P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Geophysics Dept., Stanford Univ., Stanford, CA, May 1975.
4. J. Makhoul, "A Class of All-Zero Lattice Digital Filters: Properties and Applications," to be published in IEEE Trans. Acoust., Speech, Signal Processing, 1978.
5. J. Makhoul, "Stable and Efficient Lattice Methods for Linear Prediction," IEEE Trans. Acoust., Speech, Signal Processing, pp. 423-428, Oct. 1977.
6. P.F. Fougere, "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis," J. Geophysical Res., Vol. 82, No. 7, pp. 1051-1054, March 1977.
7. R. Viswanathan and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," IEEE Trans. Acoust., Speech, Signal Processing, Vol. 23, No. 3, pp. 309-321, June 1975.
8. J. Makhoul and R. Viswanathan, "Adaptive Lattice Methods for Linear Prediction," IEEE Int. Conf. Acoustics, Speech, Signal Processing, Tulsa, OK, April 1978.
9. L.J. Griffiths, "An Adaptive Lattice Structure for Noise-Cancelling Applications," IEEE Int. Conf. Acoustics, Speech and Signal Processing, Tulsa, OK, April 1978.

10. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-23, pp. 67-72, Feb. 1975.

### Figures

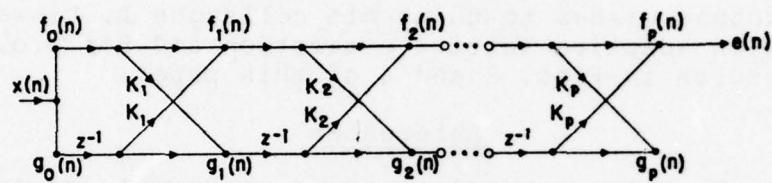


FIGURE 1. Basic all-zero lattice.

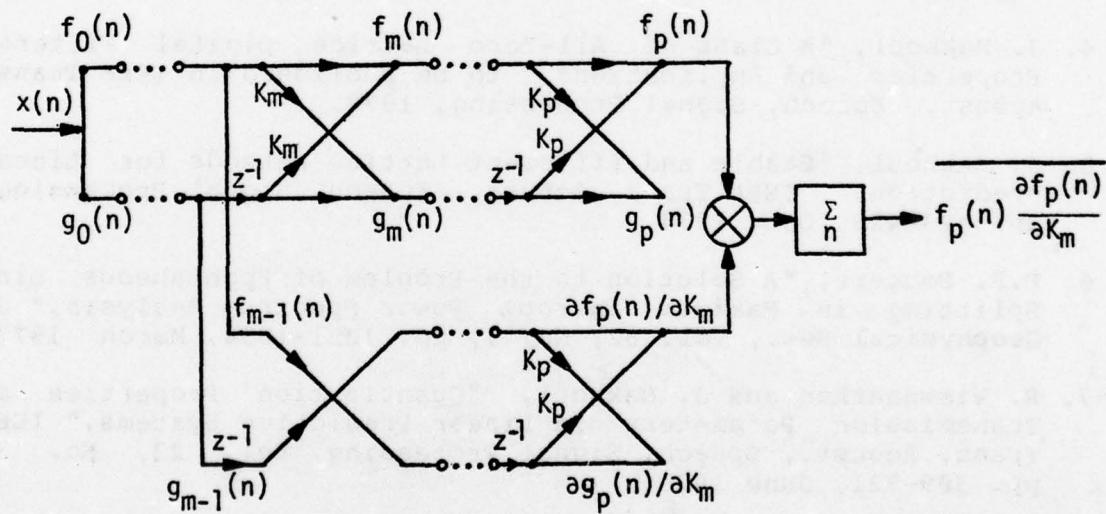


FIGURE 2. Error gradient computation using the lattice.

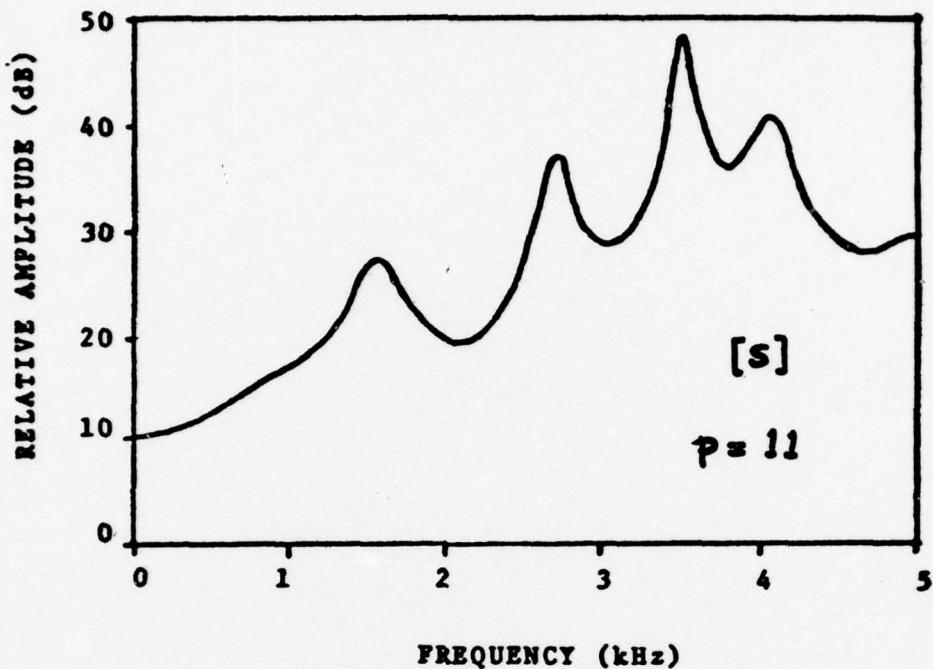


FIGURE 3. 11-pole spectrum for example 2 in text. (The spectrum corresponds to that on an [s] sound.)

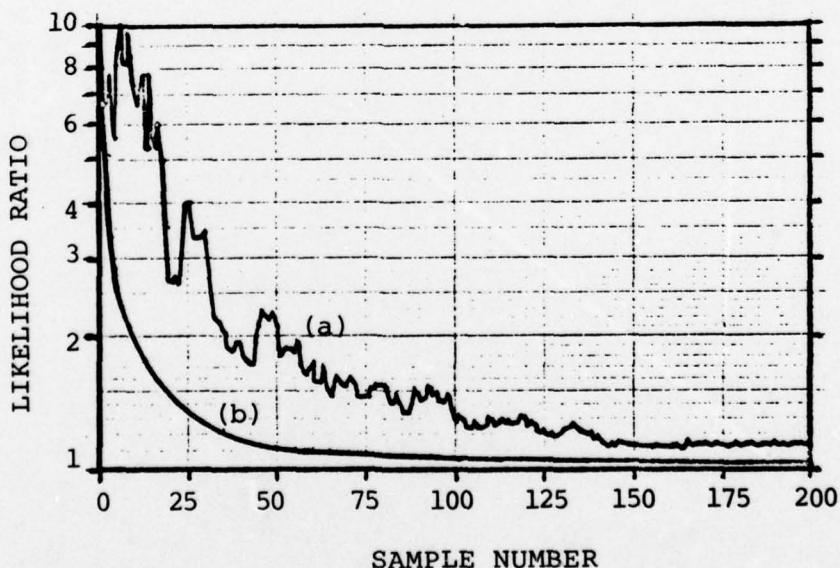


FIGURE 4. Performance of adaptive lattice as a function of sample number. (a) Plot for a single record.  
(b) Average over 15 records.

## Non-Parametric Spectrum Estimates Motivated by the Wishart Distribution

David J. Thomson

Bell Laboratories  
Whippany, New Jersey 07981

### ABSTRACT

The problems of spectrum estimation and harmonic analysis are examined using techniques suggested by properties of the complex Wishart distribution. From this viewpoint the two problems are distinct and, in the spectrum estimation case, lead to more stable estimates than do conventional nonparametric estimates. This stability is obtained by replacing the conventional estimate resulting from smoothing over frequency with a combination of several direct estimates computed using orthogonal data windows.

New results are also obtained for the harmonic analysis problem which is solved by means of a Karhunen-Loeve expansion in the frequency domain. One result of this approach is an analysis of variance test for line components in a spectrum.

### 1. Introduction

Two recurring problems in time series analysis are those of obtaining stable high resolution estimates of the spectrum when the dynamic range is large and the problem of "mixed" spectra. In the first case there is a fundamental conflict between bias and resolution which is frequently confused with a secondary trade between resolution and variance. The second problem, that of the "mixed" spectrum refers to the important practical case where a number of possibly periodic components are present together with non-deterministic noise.

While both of these problems have been extensively studied there are still very few techniques available which work well in a wide variety of situations and in recent years numerous "new" techniques and modifications of existing procedures have been introduced. Consequently one is faced with choosing one of possibly several hundred different spectrum estimates. This choice is made more difficult by misapplication of these techniques, for example the use of the "Burg algorithm" to search for periodicities, and by the lack of rigorous statistical procedures in evaluating alternative methods. Thus the classical problems of spectrum estimation, resolution, stability, hypothesis testing, and confidence statements are now more necessary than ever.

Because these problems have proved intractable using conventional spectrum estimation procedures we venture yet another method, in this case one suggested by properties of the Wishart distribution. By its definition (Anderson [1958]) the Wishart distribution is that of the estimated covariance matrix of a multivariate normal and so forms a generalization of the Chi-square distribution. However, rather than attempting to make inferences on the structure of the process by applying the properties of this distribution to the observed autocovariances of the process, the approach used here is to exploit these properties *in the frequency domain*. This procedure results in an estimation procedure which differs significantly from conventional spectrum estimates, for example a stable non-parametric estimate is obtained which does not employ conventional frequency averaging.

This viewpoint may be readily understood by regarding the spectrum estimation procedure in terms of matrix operations. We define  $\tilde{X}^t$  to be a complex row vector<sup>t</sup>

<sup>t</sup> Matrices and vectors are indicated by bold face type,  $X$ , except that  $E$  indicates the expected value operator. Fourier transforms are indicated by a "tilde",  $\tilde{X}$ , and Herméitians (conjugate transpose) by a superscript  $\dagger$  as  $X^\dagger$ . The components of vectors or matrices are indicated by the corresponding italic letter, usually subscripted,  $x_k$ , and complex conjugates by a superscript \*

$$\tilde{\mathbf{X}}' = (\tilde{x}_0^*, \tilde{x}_1^*, \tilde{x}_2^*, \dots, \tilde{x}_{m/2}^*) \quad (1.1)$$

consisting of the Fourier coefficients of the windowed data then the usual estimate of the spectrum is given by

$$\hat{\mathbf{S}} = \text{diag} < \tilde{\mathbf{X}} \tilde{\mathbf{X}}' > \quad (1.2)$$

where again  $\hat{\mathbf{S}}$  is a vector containing the spectrum estimates at the usual equidistant frequency spacings and  $\text{diag} < >$  indicates the operation of extracting the diagonal components of the matrix. One of the major disadvantages of such estimates is that except for the uni- and bivariate cases their multivariate distributions are exceedingly complex so that it is difficult to make reliable inferences from such estimates.

If however we consider the matrix

$$\tilde{\mathbf{S}} = \tilde{\mathbf{X}} \tilde{\mathbf{X}}' \quad (1.3)$$

rather than its principle diagonal it is clear that much more information is available. In particular, square submatrices taken along the diagonal of  $\tilde{\mathbf{S}}$  have a Wishart distribution. Furthermore, in regions where the spectrum is reasonably flat, the theoretical correlation matrix of the distribution is known and depends primarily on the data window. For these and other reasons which will be described in the following it is useful to study the problems of spectrum estimation and harmonic analysis in terms suggested by the Wishart distribution.

Following the introduction of some notation we consider the problem of estimating the continuous spectrum. This estimate suggests an adaptive form, section 4, computed using orthogonal data windows. The final section presents an approach to the harmonic analysis problem and a test for line components.

## 2. Basic Definitions and Notations

We consider data,  $f(t)$ , which is a realization, possibly complex, of length  $T$  from a stationary stochastic process having finite fourth moments. Such data has a *spectral representation* [ see Doob (1953) chapters X,XI ]

$$f(t) = \frac{1}{2\pi} \int e^{i\omega t} dy(\omega) \quad (2.1)$$

The data is windowed by a function  $D(t)$  and we form an estimate of the orthogonal increment process,  $dy(\omega)$ , in the continuous case by the integral

$$\tilde{x}_\omega = \int_{-T/2}^{T/2} e^{-i\omega t} f(t) D(t) dt \quad (2.2)$$

In this paper we will assume, for reasons given in Fisher (1929) that this Fourier transform,  $\tilde{x}_\omega$  has an approximately Gaussian distribution. This integral can be approximated crudely using a trapezoidal rule ( in which case it reduces to the same formula as the discrete case below ) or more accurately through the use of splines [ see Aronson(1969)]. For the discrete time case the estimate becomes:

$$\tilde{x}_\omega = \sum_{t=0}^{T-1} e^{-i\omega(t - \frac{T-1}{2})} D(t) f(t) \quad (2.3)$$

Usually the calculations are done on the discrete mesh of points  $\omega_k = 2\pi \frac{k}{m}$  where  $k = 0, 1, \dots, m/2$  when the data is real and  $k = -m/2, \dots, 0, \dots, m/2$  when the data is complex. In the following we will assume these conditions and for simplicity the notation  $\tilde{x}_{\omega_k}$  is shortened to  $\tilde{x}_k$ . It is also assumed that  $m > 2T$  always and that when high resolution is required  $m$  will be much greater than  $2T$ . This is simply the frequency domain equivalent of oversampling.

In these equations the data window,  $D$ , is properly normalized so that the common direct

estimate of spectrum

$$\hat{S}(\omega) = |\tilde{x}_\omega|^2 \quad (2.4)$$

is unbiased for white noise. In general the expected value of this spectrum estimate is given by the convolution

$$E\{\hat{S}(\omega)\} = S(\omega) * |\tilde{D}|^2 \quad (2.5)$$

which explicitly shows the dependence of the estimate on both the true spectral density function of the process and also on the data window. Further information on such estimates is available in Thomson (1977) and the references cited therein.

For reasons given in the above reference and elsewhere [ eg Harris (1978) ] we will assume that the data window,  $D$  is a spheroidal wave function, usually a *prolate spheroidal wave function*. While the properties of these functions are discussed exhaustively in Slepian, Landau, and Pollak (1961a,b, 1962) and Rhodes (1970) their most important property for the applications given here is: Of all unit energy time-limited functions they are the most concentrated, in the  $L_2$  sense, in frequency.

### 3. Estimation of the Continuous Spectrum

In this section we will be concerned primarily with the problem of forming a stable estimate of the spectrum in situations where harmonic lines do not exist. This of course is simply the classical "smoothing" problem but when the problem is approached from the Wishart matrix viewpoint the resulting "smoothers" take on the form of autoregressive decompositions applied to the Fourier transforms. To see this consider a square  $p \times p$  square complex submatrix,  $\bar{S}_k$ , taken along the main diagonal of  $\bar{S}$  about some frequency  $\omega_k$ . For example when  $p=3$  a typical submatrix is

$$\bar{S}_k = \begin{bmatrix} \tilde{x}_{k-1}\tilde{x}_{k-1}^* & \tilde{x}_k\tilde{x}_{k-1}^* & \tilde{x}_{k+1}\tilde{x}_{k-1}^* \\ \tilde{x}_{k-1}\tilde{x}_k^* & \tilde{x}_k\tilde{x}_k^* & \tilde{x}_{k+1}\tilde{x}_k^* \\ \tilde{x}_{k-1}\tilde{x}_{k+1}^* & \tilde{x}_k\tilde{x}_{k+1}^* & \tilde{x}_{k+1}\tilde{x}_{k+1}^* \end{bmatrix} \quad (3.1)$$

Clearly the submatrix  $\bar{S}_k$  is a sample covariance matrix of the form  $\tilde{X}_k \tilde{X}_k^t$  where

$$\tilde{X}_k^t = \{\tilde{x}_{k-1}^*, \tilde{x}_k^*, \tilde{x}_{k+1}^*\} \quad (3.2)$$

By definition, see Anderson (1958) chapter 7,  $\bar{S}_k$  has a Wishart distribution.

We now make the usual assumption in smoothing, that is that the spectral density is approximately constant over the width of the smoother, and with this assumption it can be seen that  $\Sigma_0$ , the population correlation matrix, is *Toeplitz*, and completely specified. Further, the population covariance matrix,  $\Sigma$ , is simply a scalar multiple of  $\Sigma_0$  with the multiplier being the unknown spectral density,  $s_k$ .

As a side point we note that if, instead of assuming that the spectrum is locally constant, we assume that it can be represented locally by a Taylor's series one obtains a population covariance matrix which can be represented by the sum

$$\Sigma = s_{k,0}\Sigma_0 + s_{k,1}\Sigma_1 + \dots \quad (3.3)$$

where  $\Sigma_0, \Sigma_1, \dots$  are known Toeplitz matrices. In this case  $s_{k,0}$  is an estimate of  $S(\omega_k)$ ,  $s_{k,1}$  of  $S'(\omega_k)$ , etc., and may be estimated by an obvious extension of the technique described below. This technique has significant implications from a smoothing viewpoint for, if one expands  $\hat{S}$  in a Fourier series, the coefficients may be estimated by simultaneously minimizing the mean squared error on the fits to both  $\hat{S}$  and  $\hat{S}'$ . Information on such fitting procedures is available in Gagayev (1957).

### 3.1. The Correlation Matrix of the Fourier Transforms

The components of the population covariance matrix,  $\sigma_{jk}$ , may be written, in the continuous time case as

$$\begin{aligned}\sigma_{jk} &= \mathbf{E} \{ x_j x_k^* \} \\ &= \int_{-T/2}^{T/2} D(x) D(u) e^{-i(\omega_j x - \omega_k u)} \mathbf{E} \{ f(x) f^*(u) \} dx du\end{aligned}\quad (3.4)$$

When the data window,  $D$ , is such that the corresponding spectral window,  $|\tilde{D}(\omega)|^2$  is *localized*, as are the windows based on *prolate spheroidal wave functions*, this integral can be simplified and becomes

$$\sigma(\zeta) = \int_{-T/2}^{T/2} D^2(t) e^{i\zeta t} dt \quad (3.5)$$

where  $\sigma(\zeta) \propto \sigma_{jk}$  and  $\zeta = \omega_j - \omega_k$ . For simplicity of notation we shall indicate  $\sigma(n\Delta\omega)$  by  $\sigma_n$  with the normalization  $\sigma_0 = 1$ . Note that, because of the time normalization factor in 2.2 the real and imaginary parts of  $\tilde{x}_j$  are uncorrelated so that  $\sigma$  describes the autocorrelation of either.

### 3.2. Spectrum Estimates from the Wishart Distribution

Using the above notations the Wishart probability density function is given by

$$\frac{|\bar{\Sigma}_k|^{1/2(n-p-1)} \exp(-1/2 \operatorname{tr}\{\Sigma^{-1} \bar{\Sigma}_k\})}{2^{1/2np} \pi^{np(p-1)} |\Sigma|^{1/2n} \prod_{l=1}^p \Gamma(1/2(n+1-l))} \quad (3.6)$$

in which  $n$  represents the number of terms added together to form each element of the sample covariance matrix and, as above, the covariance matrix is  $p \times p$ . Thus in the single sample case  $n=2$  because of the real and imaginary components. Examining the log-likelihood function one finds

$$L = -1/2 \operatorname{tr}\{\Sigma^{-1} \bar{\Sigma}_k\} - 1/2 n \ln |\Sigma| + \text{terms independent of } \Sigma \quad (3.7)$$

Now  $\operatorname{tr}\{\Sigma^{-1} \bar{\Sigma}_k\}$  is given by  $\frac{1}{s_k} \operatorname{tr}\{\Sigma_0^{-1} \bar{\Sigma}_k\}$  and  $|\Sigma|$  is  $s_k^n |\Sigma_0|$  so that subject to the above conditions the maximum-likelihood estimate of  $s_k$  is given by

$$\hat{s}_k = \frac{1}{np} \operatorname{tr}\{\Sigma_0^{-1} \bar{\Sigma}_k\} \quad (3.8)$$

Under the smoothness and localization assumptions given above  $\Sigma_0$  is Toeplitz so that its inverse may be found simply and the spectrum estimate given in 3.8 written as a weighted sum of squares. This is most easily seen by writing  $\Sigma_0$  in the form, Burg(1972)

$$\Sigma_0^{-1} = \mathbf{A} \Gamma^{-1} \mathbf{A}^T \quad (3.9)$$

where  $\mathbf{A}$  is a lower triangular matrix consisting of the prediction error filters computed from the correlations,  $\sigma(k\Delta\omega)$ , of the estimated orthogonal increment process. Similarly  $\Gamma$  is a diagonal matrix of the corresponding frequency domain prediction errors,  $\Gamma = \operatorname{diag}(1, \gamma_1, \dots, \gamma_{p-1})$ . Thus the quantity  $\operatorname{tr}\{\Sigma_0^{-1} \bar{\Sigma}_k\}$  may be written as

$$\operatorname{tr}\{\mathbf{A} \Gamma^{-1} \mathbf{A}^T \mathbf{X}_k \mathbf{X}_k^*\} \quad (3.10)$$

which, by the properties of the trace operator becomes

$$\operatorname{tr}\{\mathbf{X}_k^* \mathbf{A} \Gamma^{-1} \mathbf{A}^T \mathbf{X}_k\} \quad (3.11)$$

Since  $\Gamma$  is diagonal this can be written as a sum of squares so that

$$\hat{s}_k = \frac{1}{np} \sum_{j=0}^{p-1} \frac{1}{\gamma_j^2} \left| \sum_{h=0}^j \alpha_h^{(j)} x_{k+[p/2]-j+h} \right|^2 \quad (3.12)$$

This form is interesting in its implications as it basically consists of a sum of direct spectrum estimates with different data windows. The first of these windows is the one selected, i.e.  $D$ , the second is almost its first derivative, and so on. Also the correlations between the estimates in the above sum are, in the case of locally flat spectra, low and may be described in terms of the partial correlations of the data window.

Clearly this estimate is positive. Also, in common with direct and indirect estimates and in contradistinction to most parametric estimates, the estimate corresponding to several independent data sets is simply the average of the individual estimates.

#### 4. Direct Estimates Based on Orthogonal Data Windows

However, since the title of this paper is "Estimates Motivated by the Wishart Distribution", we make the obvious generalization and replace the approximately orthogonal data windows with a strictly orthonormal set of data windows. Specifically we choose to use the first  $p$  prolate spheroidal wave functions or their generalizations and form the  $p$  estimates of spectra

$$\hat{S}_k^{(j)} = \int_{-T/2}^{T/2} e^{-i\omega_k t} f(t) D_j(t) dt \quad j=0, 1, \dots, p-1 \quad (4.1)$$

where the data windows  $D_j(t)$  are proportional to the  $j^{\text{th}}$  prolate spheroidal wave function  $S_{0,j}(c, 2t/T)$  and are normalized by the condition

$$\int_{-T/2}^{T/2} D_j(t)^2 dt = 1 \quad (4.2)$$

In these windows the parameter  $c = \Omega T/2$  represents the time bandwidth product and the windows have the property that the integral of the spectral window over  $(-\Omega, \Omega)$  is given by

$$\frac{1}{2\pi} \int_{-\Omega}^{\Omega} |\tilde{D}_j(\omega)|^2 d\omega = 1 - \lambda_j(c) \quad (4.3)$$

where  $\lambda_j(c)$  is the  $j^{\text{th}}$  eigenvalue of the integral equation defining the functions [ Slepian & Sonnenblick (1965) ].

Now, by using arguments given in detail in Thomson (1977a) the *broad band bias* of the  $j^{\text{th}}$  estimate is bounded by a term proportional to  $1 - \lambda_j(c)$ . These eigenvalues have the property that

$$1 > \lambda_0(c) > \lambda_1(c) > \dots \geq 0 \quad (4.4)$$

so that the bias properties of the estimates have a natural rank ordering. Obviously the estimate having the minimum bias is that obtained using the  $0^{\text{th}}$  function, and choosing  $c = 4\pi$  as an example  $1 - \lambda_0 \approx 3 \cdot 10^{-10}$ . This window†, discussed at length in Thomson *et al* (1976) has superb bias properties. A second important property of these eigenfunctions is that there are approximately  $2c/\pi$  "large" eigenvalues after which they rapidly decay. This number sets an upper bound on the parameter  $p$ , and for the  $4\pi$  window the use of the first 4 or 5 functions is reasonable so that in these terms the estimate 3.12 above is simply

$$\bar{S}_k = \frac{1}{p} \sum_{j=0}^{p-1} \hat{S}_k^{(j)} \quad (4.5)$$

In regions where the spectrum may be regarded as "flat" over the frequency range  $(\omega_k - 2c/T, \omega_k + 2c/T)$  this estimate is distributed as a  $\chi^2_{2p}$ . As an example for  $c = 4\pi$  and  $p = 5$  this technique results in 10 degrees of freedom, better than the 6.7 dofs obtainable by frequency averaging the  $0^{\text{th}}$  estimate over the same bandwidth and is reasonably efficient relative to the 16 dofs optimistically available.

† Named by the authors colleagues the "Thomson" window.

To see the behaviour of these estimates better figure 1 shows the first four prolate spheroidal wave functions in the time domain (and since they are their own Fourier transforms also in the frequency domain). Figure 2 shows the corresponding spectral windows as a function of frequency normalized to units of  $1/T$ . Figure 3 shows the spectral window obtained by averaging the estimates according to 4.5 as a function of  $\rho$ , and it can be seen that the spectral window approaches an "ideal" rectangular form. However, since the eigenvalues,  $\lambda_p$ , decrease, the bias of this simple estimate is increased.

The fact that these eigenvalues have a natural ordering suggests a further generalization in the form of an adaptive estimate. If one now computes the regression coefficient between the logarithms of the different spectrum estimates and  $\ln(1 - \lambda_k)$ ,

$$C_k = \sum_{j=0}^{p-1} \Lambda_j \left[ \ln\{\hat{s}_k^{(j)}\} - \overline{\ln\{\hat{s}_k^{(\cdot)}\}} \right] \quad (4.6)$$

where  $\overline{\ln\{\hat{s}_k^{(\cdot)}\}}$  indicates the average over the logarithms of the  $p$  spectrum estimates, and

$$\Lambda_j = \frac{\ln(1 - \lambda_j) - \frac{1}{p} \sum_{h=0}^{p-1} \ln(1 - \lambda_h)}{\left[ \sum_{l=0}^{p-1} \{ \ln(1 - \lambda_l) - \frac{1}{p} \sum_{h=0}^{p-1} \ln(1 - \lambda_h) \}^2 \right]^{\frac{1}{2}}} \quad (4.7)$$

one has a powerful test for violation of the "localized estimate" hypothesis. Under this hypothesis, that is that the differences between the  $p$  spectrum estimates reflect only sampling variation from within the "local" frequency region of  $\pm 2c/T$  and that "broad band" bias is insignificant, one has  $E\{C_k\}=0$  and  $\text{var}\{C_k\} \approx \pi^2/6\Delta$  where  $\Delta$  is the term in the denominator of 4.7. Under the alternative hypothesis, that is that the estimates are being influenced by "broad band" bias more than by local conditions, the dependence of this bias on the eigenvalues will be seen as large positive values of the regression coefficient  $C_k$ . In such cases the estimate can be made adaptive by using the following procedure which tends to ignore the more biased estimates.

Following computation of the regression coefficient,  $C_k$ , the adaptive procedure is initiated by setting the unnormalized spectrum estimate

$$\hat{U}_k^{(1)} = \hat{s}_k^{(0)} + \hat{s}_k^{(1)} \quad (4.8)$$

and a cumulative weight

$$W_k^{(1)} = 2 \quad (4.9)$$

after which steps 4.10-4.13 are done for  $j = 2, \dots, p-1$ . The first step of this iterative process is to compute an estimate of the broad band bias for the  $j^{\text{th}}$  estimate

$$\hat{b}_k^{(j)} = \exp \left[ \overline{\ln\{\hat{s}_k^{(\cdot)}\}} + C_k \ln(1 - \lambda_j) \right] \quad (4.10)$$

which is used with the current estimate of the spectrum to compute a weight

$$w_k^{(j)} = \frac{1}{1 + \left[ \frac{b_k^{(j)}}{U_k^{(j-1)} / W_k^{(j-1)}} \right]^2} \quad (4.11)$$

next, the unnormalized sum and cumulative weight are updated

$$\hat{U}_k^{(j)} = \hat{U}_k^{(j-1)} + w_k^{(j)} \hat{s}_k^{(j)} \quad (4.12)$$

$$W_k^{(j)} = W_k^{(j-1)} + w_k^{(j)} \quad (4.13)$$

and the final estimate given by

$$\bar{S}_k = U_k^{(p-1)} / W_k^{(p-1)} \quad (4.14)$$

As an example of a situation where such a procedure is useful it has been hypothesized [Gans (1972)] that the temporal fading characteristics of a particular mobile communications channel may be approximately described in terms of a stationary process with a Bessel autocorrelation function. This autocorrelation function corresponds to a band-limited spectrum where the bandlimit is given by the mobile's doppler frequency. To check the accuracy of such a model it is necessary to have confidence in the processing algorithm and data corresponding to this model was generated using a Karhunen-Loeve expansion. To speed up the simulations a frequency translated replication of the process was added to the hypothesized low pass process so that the behaviour across three bandedges is observable.

The upper frame of figure 4 shows an example of 127 such data points (1 second) in the time domain. The lower frame shows the spectrum estimates obtained using the first five prolate spheroidal wave functions with  $c = 4\pi$  as data windows and it may be seen that, as expected, those corresponding to the functions of order 3 and 4 are badly biased near the band edges. The adaptive estimate is demonstrated in figure 5 with the upper frame showing the regression coefficient standardized by its theoretical variance. Clearly it is very effective as a bias indicator and, as can be seen in the lower frame the adaptive spectrum estimate has the stability of a smoothed estimate in regions where bias on the higher order functions is low and the low bias properties of the zero order function elsewhere.

##### 5. Harmonic Analysis and a Test of Significance

As mentioned in the introduction one of the outstanding problems in spectrum estimation is that of "mixed" spectra, that is the case where the data contains a number of periodic signals in addition to the usual purely nondeterministic component [ Priestley (1962) ]. The first problem in this situation is to decide if, in fact a given peak in an estimated spectrum is a result of a line component or if it is simply extreme sampling variation. For this purpose the usual approach is to compare the level of the peak with that of the local continuous component and apply an  $F$  test with 2 and "several" degrees of freedom.

In principle this situation can be handled by expanding the process in the time domain using a Karhunen-Loeve expansion with a non-zero mean value function consisting of a trigonometric series and solving the resulting likelihood equations for the coefficients of the series. This approach however presupposes a knowledge of both the number and frequencies of any existing lines, the covariance function of the non-deterministic component, and also is impractical from a computational viewpoint for all except very short series.

As before we make the assumption that the portion of the spectrum due to non-deterministic components of the spectrum is slowly varying so that the matrix  $\bar{S}_k$  is approximately Wishart with a Toeplitz covariance matrix. The presence of the line component however implies that  $\bar{S}_k$  is non-central Wishart so that [ see Muirhead (1978) ] direct application of likelihood procedures becomes exceedingly complex. This complexity may be circumvented by using an eigenvalue decomposition for  $\Sigma^{-1}$  so that the test which we propose uses the correlation properties of the windowed and transformed data and, in fact, is implemented using a discrete Karhunen-Loeve expansion in the frequency domain. This approach avoids the outstanding problems with the time domain method: the covariance function of the estimated orthogonal increment process is known, the shape of the mean value function is known, and finally the use of good windows restricts the frequency range which must be examined.

In this application use of the Karhunen-Loeve expansion is dictated by numerical considerations. The transformed data is a band limited function so that direct matrix inversion (for example as required in the formally equivalent generalized least squares technique) is unstable. The use of the truncated Karhunen-Loeve expansion is sufficient for the estimation procedure while retaining numerical accuracy.

As above we consider a purely non-deterministic time series but now with the addition of a line component at a frequency  $\omega_0$ . In this case the  $\tilde{x}(\omega)$  has a mean value function  $\mu \tilde{D}(\omega - \omega_0)$  where  $\mu$  is proportional to the amplitude of the periodic component. Again on the assumption that the continuous component of the spectrum is slowly varying in the vicinity

of  $\omega_0$  one can write the likelihood in terms of a Karhunen-Loeve expansion about some center frequency  $\omega_k$  as

$$L(k) = (\tilde{\mathbf{X}}_k - \mu \tilde{\mathbf{D}})^t \Psi \Lambda^{-1} \Psi^t (\tilde{\mathbf{X}}_k - \mu \tilde{\mathbf{D}}) \quad (5.1)$$

In this equation the eigenvalues and eigenfunctions are defined by the matrix equation

$$\Lambda \Psi^t = \Sigma \Psi^t \quad (5.2)$$

with the eigenfunctions normalized by  $\Psi^t \Psi = I$ . Note however, that because of the numerical problems referred to above, that only the eigenvectors corresponding to the first  $h$  largest eigenvalues are retained. Thus  $\Psi^t$  is a  $h \times p$  matrix while  $\Lambda$  is an  $h \times h$  diagonal matrix of these eigenvalues. The vector describing the mean value function,  $\tilde{\mathbf{D}}$  is defined as

$$\tilde{\mathbf{D}}^r = (\tilde{D}(-\frac{p-1}{2}\Delta\omega), \tilde{D}(-\frac{p-3}{2}\Delta\omega), \dots, \tilde{D}(+\frac{p-1}{2}\Delta\omega)) \quad (5.3)$$

Maximizing the likelihood the estimate,  $\hat{\mu}_k$ , of a possible line component amplitude is given by the scalar equation

$$\tilde{\mathbf{D}}^t \Psi \Lambda^{-1} \Psi^t \tilde{\mathbf{X}}_k = \hat{\mu}_k \tilde{\mathbf{D}}^t \Psi \Lambda^{-1} \Psi^t \tilde{\mathbf{D}} \quad (5.4)$$

Under either hypothesis estimating a possible mean value function accounts for 2 degrees of freedom so that

$$\hat{\sigma}_k^{(r)2} = (\tilde{\mathbf{X}}_k - \hat{\mu}_k \tilde{\mathbf{D}})^t \Psi \Lambda^{-1} \Psi^t (\tilde{\mathbf{X}}_k - \hat{\mu}_k \tilde{\mathbf{D}}) \quad (5.5)$$

is distributed proportionally to a  $\chi_{2h-2}^2$ . Similarly, under the null hypothesis,

$$\hat{\sigma}_k^{(\mu)2} = |\hat{\mu}_k|^2 \tilde{\mathbf{D}}^t \Psi \Lambda^{-1} \Psi^t \tilde{\mathbf{D}} \quad (5.6)$$

will be proportional to a  $\chi_2^2$ . Since these two variance estimates are independent their ratio,

$$F_k = \frac{\hat{\sigma}_k^{(r)2}}{\hat{\sigma}_k^{(\mu)2}} \quad (5.7)$$

will have, under the null hypothesis, an  $F$  distribution with 2 and  $2h-2$  degrees of freedom which may be tested for significance by standard methods.

In practice the test for line components has been modified so that a search procedure is used. The first stage of this procedure is to tag all local maxima of the estimated spectrum. In the second stage the  $F$  statistic is computed over a mesh of points about these local maxima and that frequency which has the maximum  $F$  ratio identified as "the frequency" of a possible line component. Also the summation is modified so that only even eigenfunctions are included; this has the effect of increasing the sensitivity of the test and also of making it insensitive to slope in the spectrum. The final step in the procedure is, when a significant line component is detected, to first remove the estimated mean value function from the Fourier transform, and second to modify the resulting spectrum by replacing the power thus removed at the estimated line frequency.

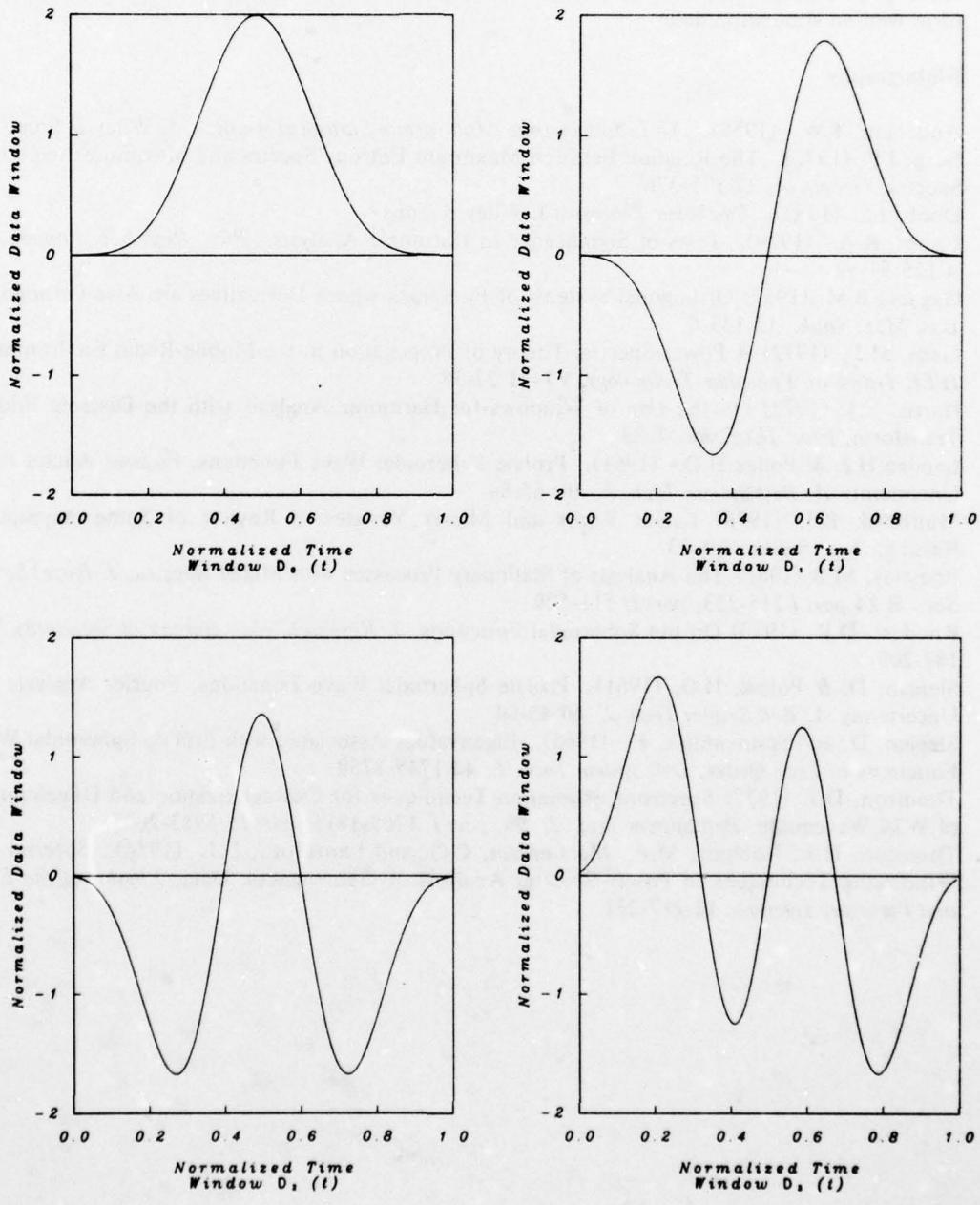
As a test of this theory data similar to that used in the preceding example with the addition of an  $AR-5$  base component and three low level non-harmonic sinusoids was generated. Figure 6 shows a typical time domain realization (127 data points) and the unsmoothed direct estimate computed using the  $4\pi$  prolate window. Also plotted is the "marker" sequence showing the peaks to be examined. The results of this examination are shown in figure 7. The upper grid gives values of  $F$  at the selected peaks while the lower grid shows the spectrum estimate with the "redistribution" of power. Limited experience with this procedure has shown that it is accurate both with respect to correctly classifying peaks as lines or sampling variation and also for correctly estimating the frequencies of peaks with typical accuracies being better than  $1/4T$ .

## 6. Summary and Conclusions

Using frequency domain characteristics of tapered and transformed samples of a stationary time series suggested by the Wishart distribution two new techniques for time series analysis have been described. The first of these is well adapted to estimating the continuous component of a spectral density function and combines high resolution with stability. The second technique gives a simple and accurate test for line components in a spectrum. Both techniques perform well on simulation data.

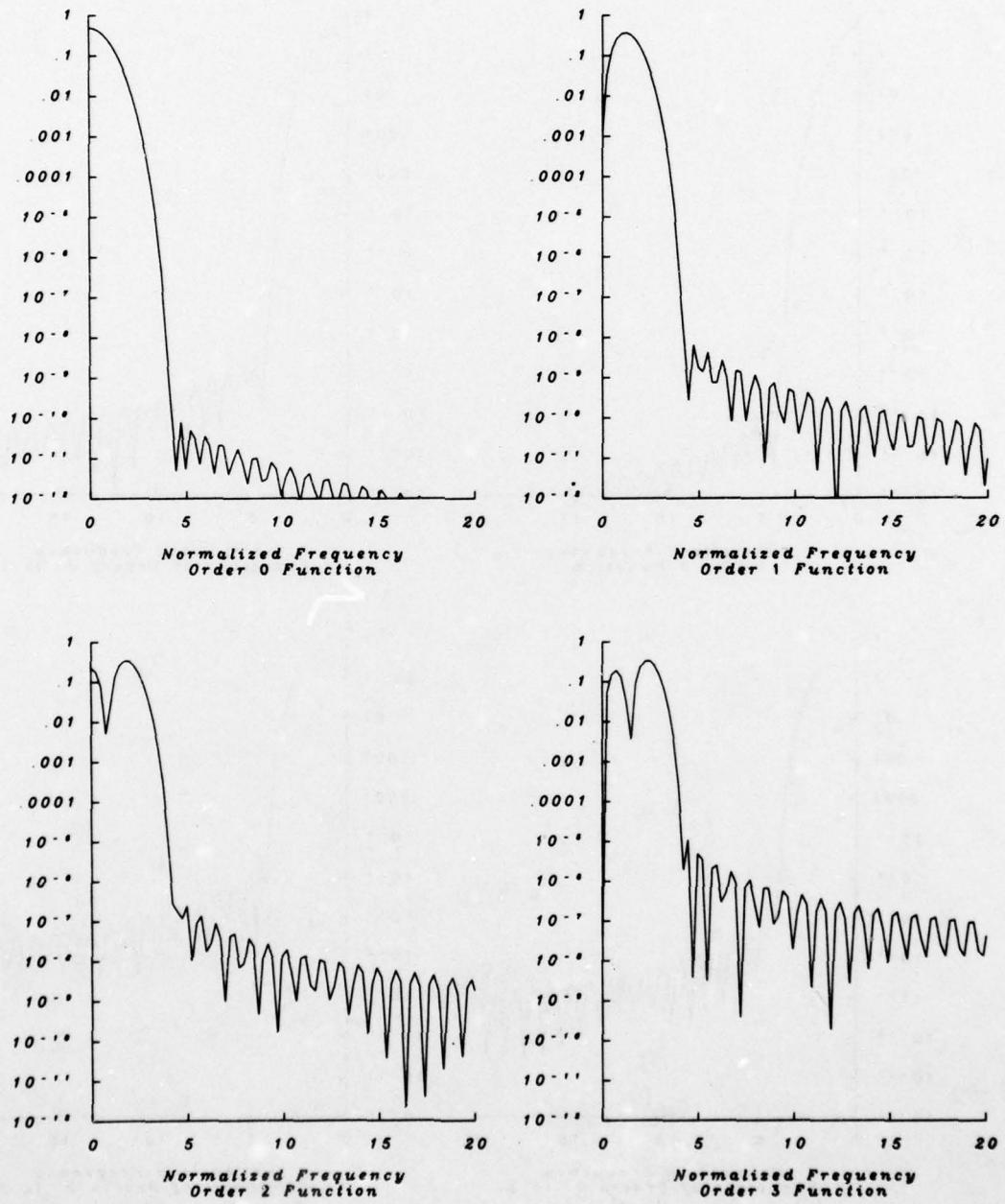
## Bibliography

- Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*, J. Wiley & Sons  
Burg, J.P. (1972). The Relation Between Maximum Entropy Spectra and Maximum Likelihood Spectra, *Geophysics* **37** 375-376  
Doob, J.L. (1953). *Stochastic Processes* J. Wiley & Sons  
Fisher, R.A. (1929). Tests of Significance in Harmonic Analysis. *Proc. Roy. Soc. London Ser. A* **125** 54-59  
Gagaev, B.M. (1957) Orthogonal Systems of Functions whose Derivatives are Also Orthogonal, *Upsi Mat. Nauk.* **12** 133-6  
Gans, M.J. (1972) A Power-Spectral Theory of Propogation in the Mobile-Radio Environment, *IEEE Trans. on Vehicular Technology*, **VT-21** 27-38  
Harris, F.J. (1978) On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform, *Proc. IEEE*, **66** 51-83  
Landau,H.J. & Pollak,H.O., (1961). Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-II, *Bell System Tech. J.* **40**, 65-84  
Muirhead, R.J. (1978) Latent Roots and Matrix Variates: a Review of Some Asymptotic Results, *Ann. Statist.* **6** 5-33  
Priestley, M.B (1962) The Analysis of Stationary Processes with Mixed Spectra, *J. Royal Statist. Soc. B* **24 part I** 215-233, *part II* 511-529  
Rhodes, D.R. (1970) On the Spheroidal Functions, *J. Research, Nat. Bureau of Standards*, **74B** 187-209  
Slepian, D. & Pollak, H.O. (1961). Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty -I, *Bell System Tech. J.* **40** 43-64  
Slepian, D. and Sonnenblick, E. (1965). Eigenvalues Associated with Prolate Spheroidal Wave Functions of Zero Order, *Bell System Tech. J.* **44** 1745-1759  
Thomson, D.J. (1977) Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide, *Bell System Tech. J.* **56**, *part I*, 1769-1815, *part II*, 1983-2005.  
Thomson, D.J., Robbins, M.F., MacLennan, C.G. and Lanzerotti, L.J. (1976). Spectral and Windowing Techniques in Power Spectral Analysis of Geomagnetic Data, *Physics of the Earth and Planetary Interiors*, **12** 217-231



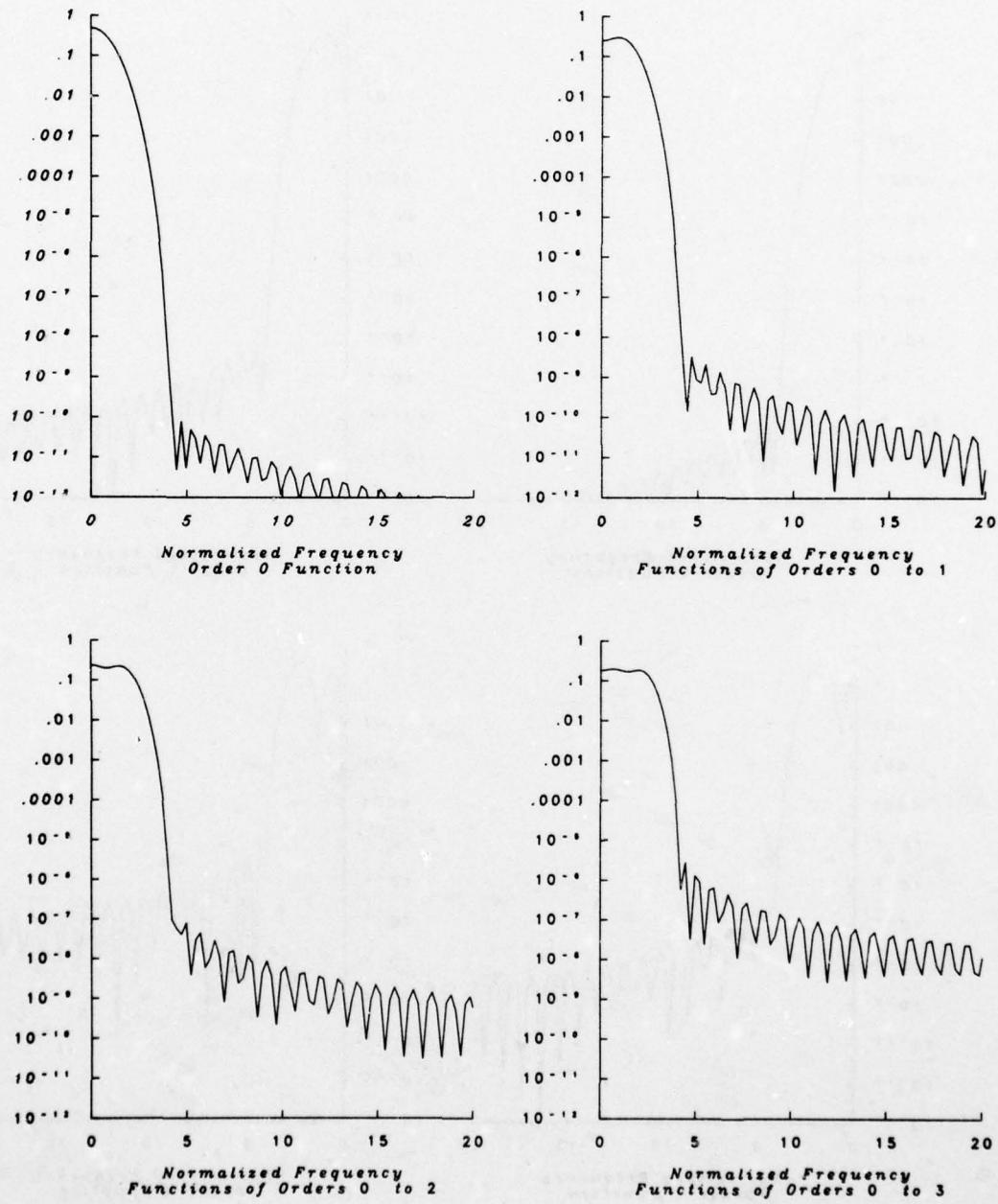
Prolate  $4\pi$  Data Windows

Figure 1



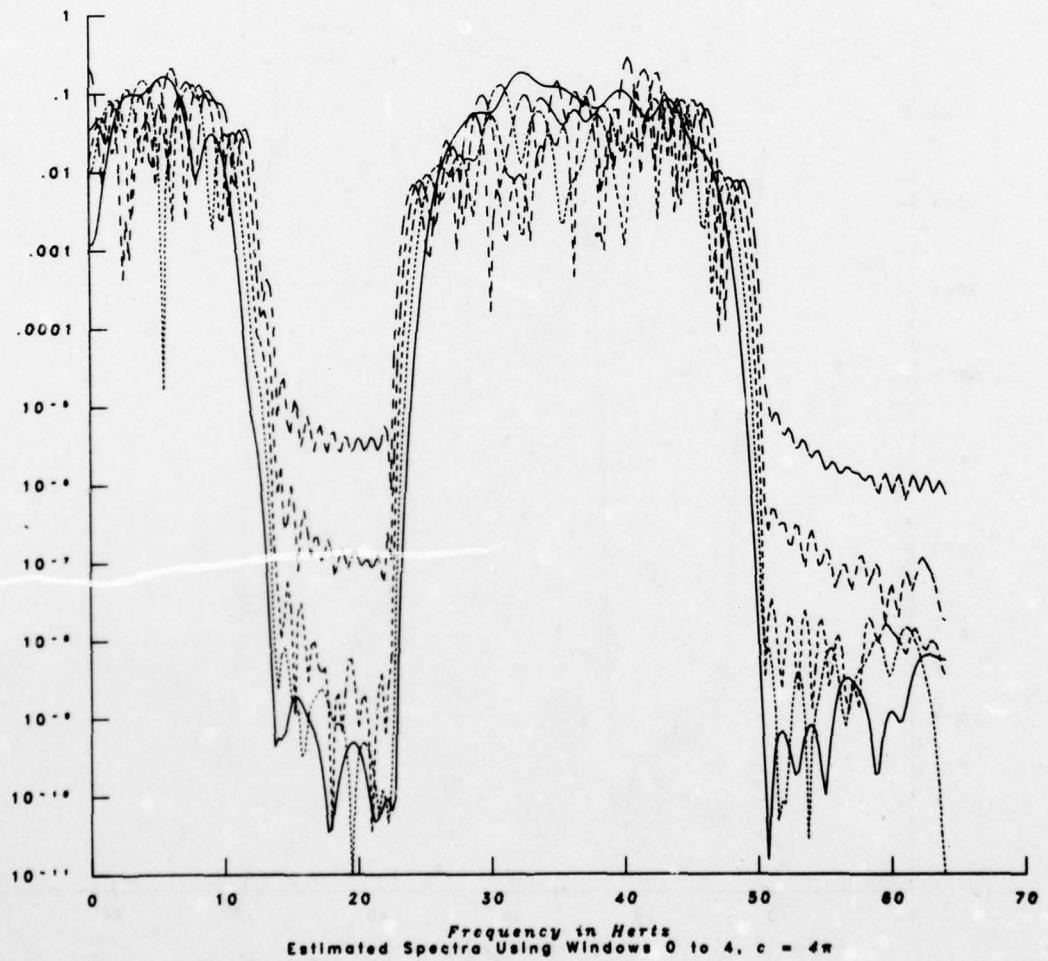
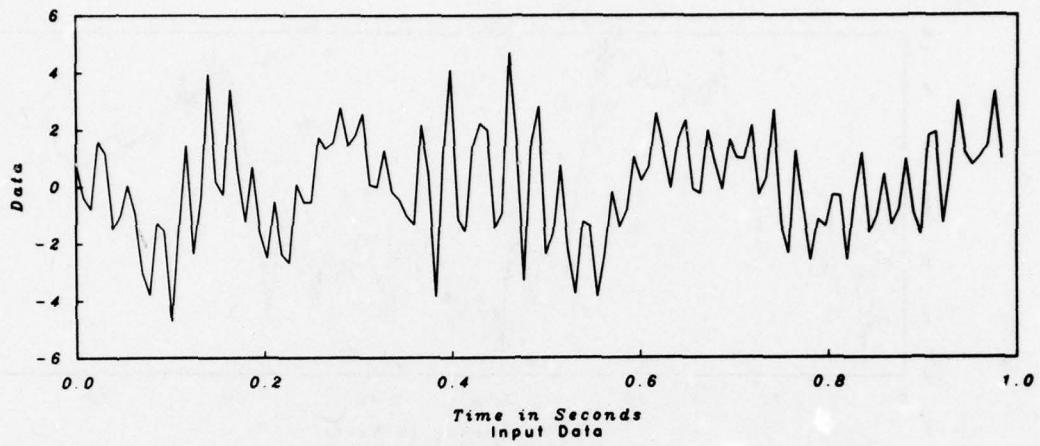
Prolate  $4\pi$  Spectral Windows

Figure 2



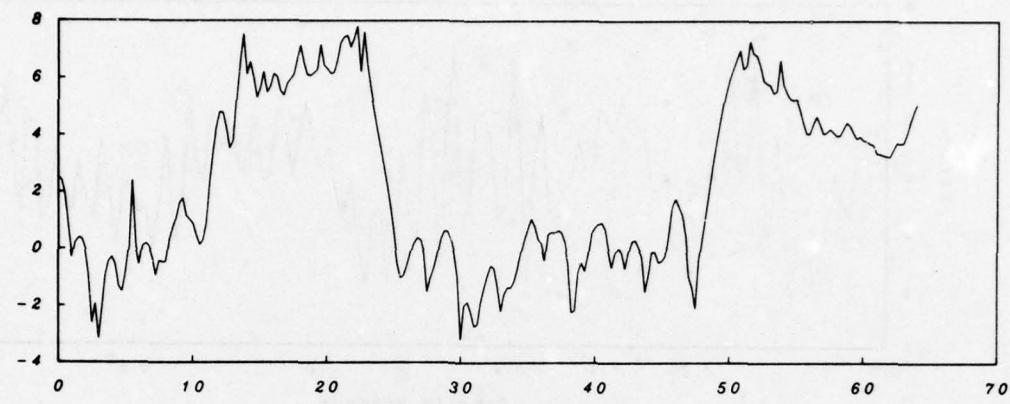
Averaged Prolate  $4\pi$  Data Windows

Figure 3

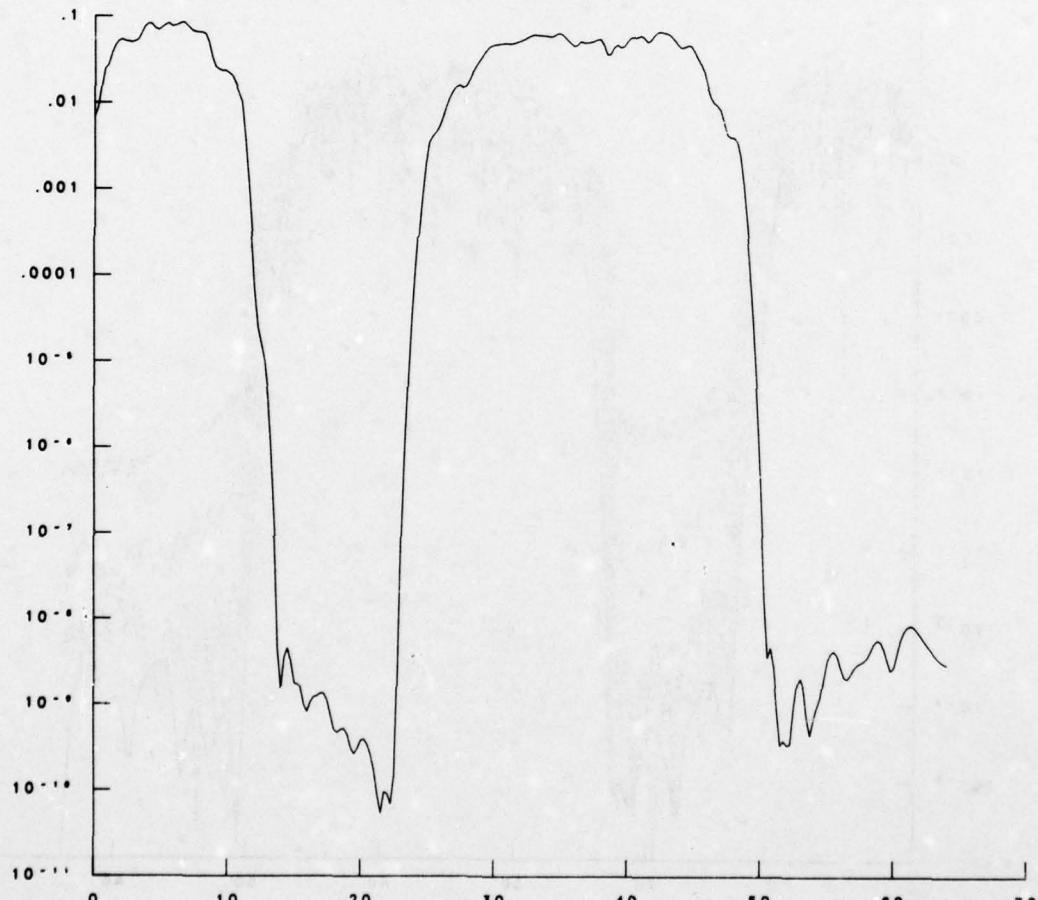


Simulation Data,  $(1+\cos)x/(x)$  Covariance

Figure 4



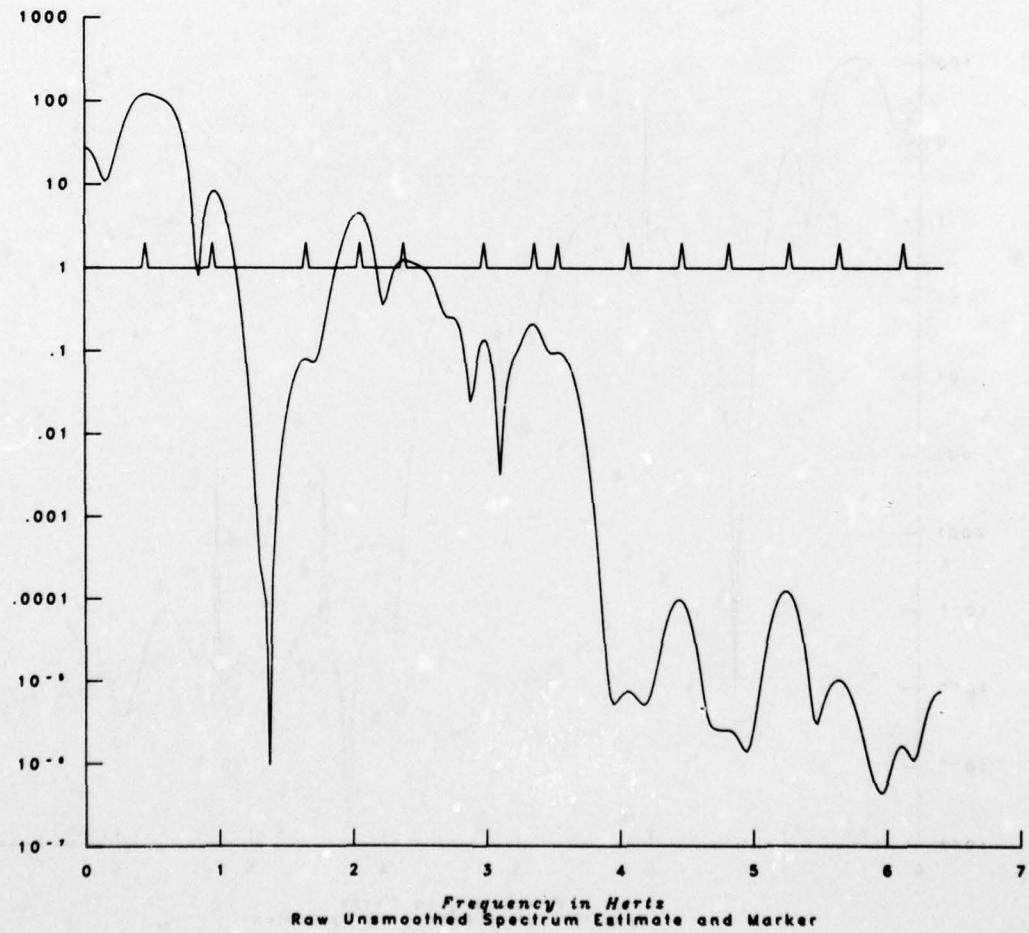
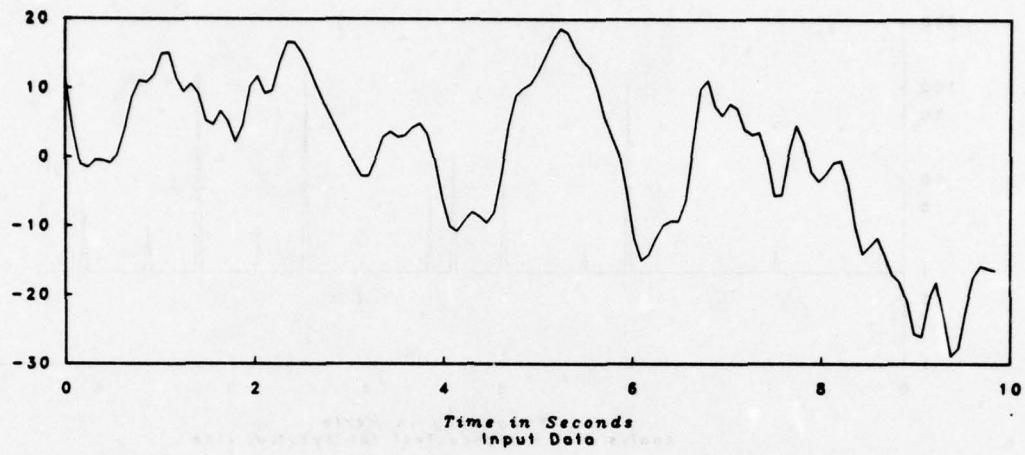
Frequency in Hertz  
Bias Test Criterion C<sub>0</sub>



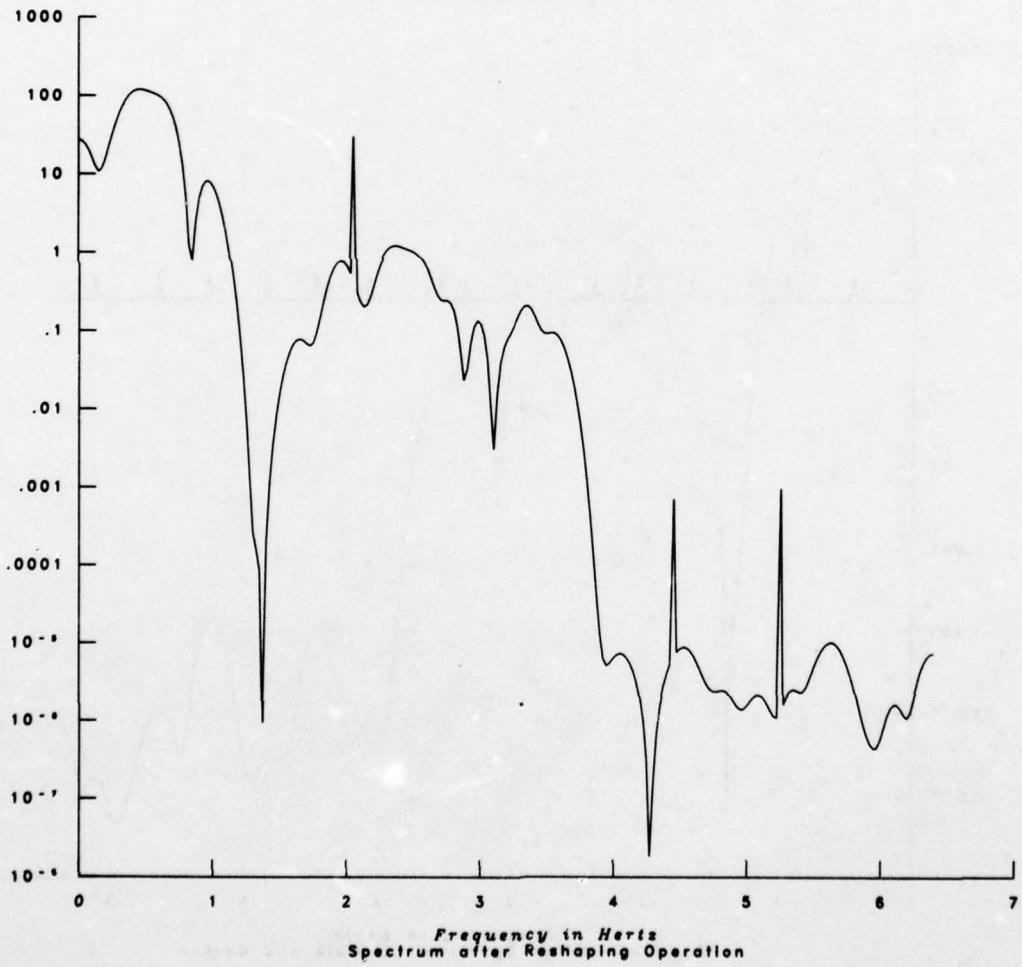
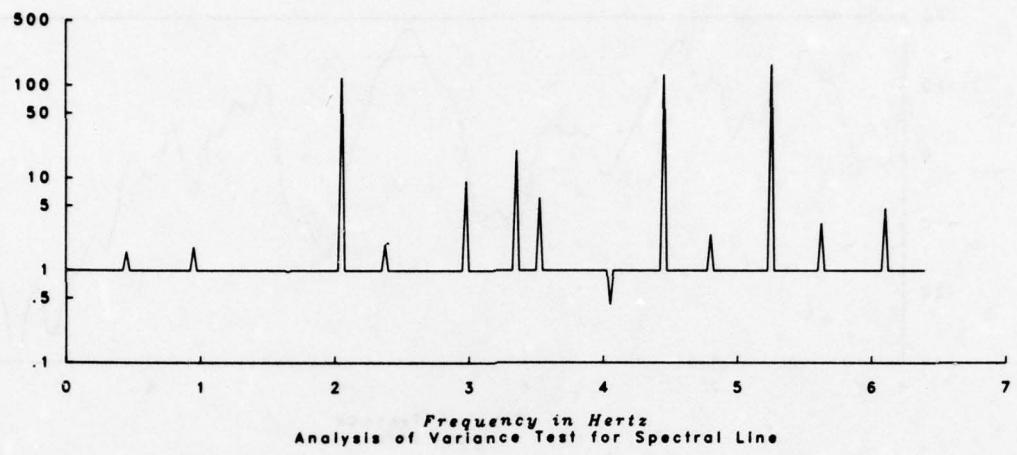
Frequency in Hertz  
Adaptive Spectrum Estimate with 5 Data Windows

Simulation Data,  $(1+\cos)x/(x)$  Covariance

Figure 5



Composite Simulation, Bessel + AR-5 + Lines  
Figure 6



Composite Simulation, Bessel + AR-5 + Lines  
Figure 7

AIR VEHICLE DETECTION  
USING ADVANCED SPECTRAL ESTIMATION TECHNIQUES

PHILIP G. TOMLINSON  
GUY A. ACKERSON

Decision-Science Applications, Inc.  
1500 Wilson Boulevard  
Arlington, Virginia 22209

Introduction

This paper concerns itself with the class of radar surveillance applications generally referred to as the "look-down" radar problem. A typical example is detection of low flying air vehicles using airborne or spaceborne radars. This paper addresses the use of advanced spectral estimation techniques (as compared with the conventional Fourier transform) to distinguish between signals from targets and those from background clutter.

This paper should be viewed as a status report in as much as the investigation was only recently initiated. Of what follows, approximately one third is devoted to an exposition of the problem, one third to the spectral estimation techniques considered and one third to preliminary results. These results are promising and it is hoped that this report will interest others in the problem and motivate further research.

The Look-Down Radar Problem

Horizon limits and local terrain features can severely restrict a land-based radar's capability to detect low flying targets. This limitation has generated much interest in airborne or spaceborne radar platforms which have a much greater field-of-view. There are difficulties which must be overcome. Background clutter is a major one. In most cases the earth intersects the radar beam pattern and much of the return from the earth (ground clutter) can not be separated from the target using the radar's range and angle resolution capability (see Fig. 1). A typical space radar viewing low flying targets against the earth's background will encounter 50-60 dBsm of clutter radar cross section (RCS).

---

This work was supported, in part, by Rome Air Development Center under Contract F30602-77-C-0231 and F30602-75-C-0122.

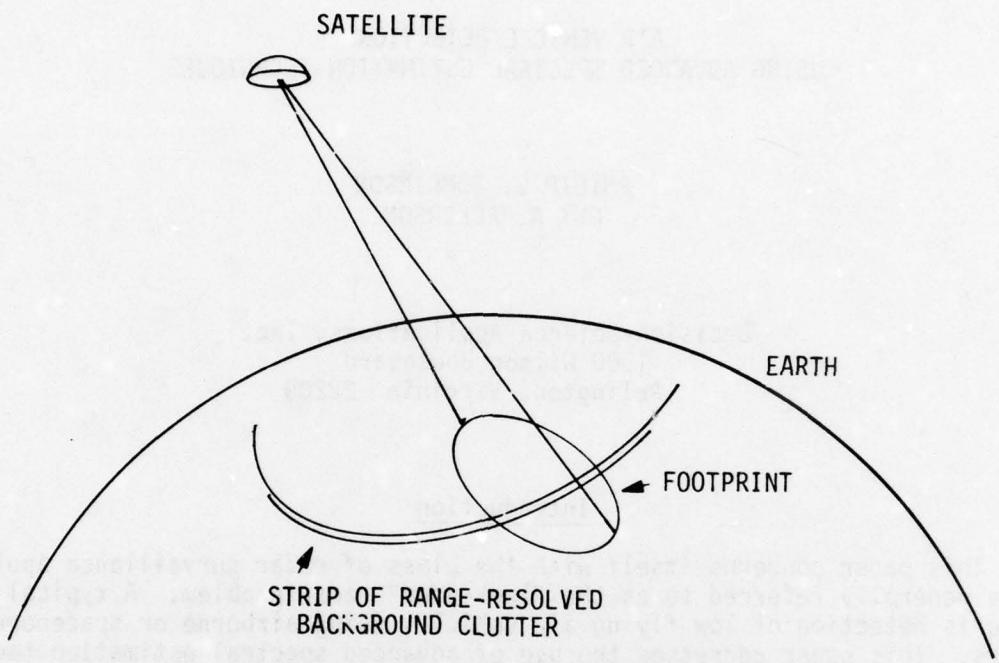


FIGURE 1. Schematic of Spaceborne Look-Down Radar

The traditional approach to this problem is to separate the targets from this clutter by exploiting the Doppler shifted frequency of a return from a moving target. The spectral characteristics of the radar return and the ability to measure this spectrum determine the capability of the look-down radar.

Processing the radar return is an obvious area where advanced spectral estimation techniques may prove useful. To be an effective estimation procedure in this application, the method must provide a high resolution spectral estimate, but must also possess several properties not always present in the newer spectral estimation techniques. The techniques must accommodate a noisy, multiple target environment. The technique must be effective when there is little or no apriori knowledge about the target frequency. The range component of the target's velocity determines the Doppler shift and variations in speed and direction of flight cause large variations in range-rate. The method must relatively faithfully reproduce the amplitudes of the frequency components over a large dynamic range. Unwanted signals can easily exceed the desired target signals by factors of a 1000. If a relatively linear response is not maintained over this dynamic range, then various threshold tests applied to the spectrum to separate targets from noise and clutter will not be effective.

Figure 2 shows a typical clutter spectrum with typical targets. This spectrum is characterized by a strong central response coming from clutter illuminated by the radar mainbeam. Additional ground clutter components come in through the antenna sidelobes which illuminate a large portion of the earth's surface. In addition there are imperfections in the radar transmitter and receiver hardware which introduce unwanted signals and a noise-like clutter contribution. The result is a relatively white clutter floor which extends over the frequency interval of interest and is proportional to the total clutter signal.

The separation of the targets from the clutter requires a relatively narrow-band Doppler filter. This in turn requires a relatively long sample of the signal. The transmission and reception of this long time sample occupies the radars time and is not compatible with high search rates and limits radar capability. Reduced dwell times result in lower signal-to-interference ratios and poorer detection statistics. Any spectral estimation procedure which can improve the effective signal-to-interference ratio or the frequency resolution capability of the radar while limiting the dwell time would be extremely valuable.

With this in mind, consider two example problems where ultimate performance is restricted by a dwell time constraint. The first example concerns itself with targets such as "target a" in Fig. 2. Here the viewing geometry and target's velocity are such that it competes only with the noise-like clutter floor. A pulse-burst waveform is transmitted. Traditionally the receiver processes the return using a discrete Fourier transform and an amplitude taper (i.e. window). This target is isolated from the major

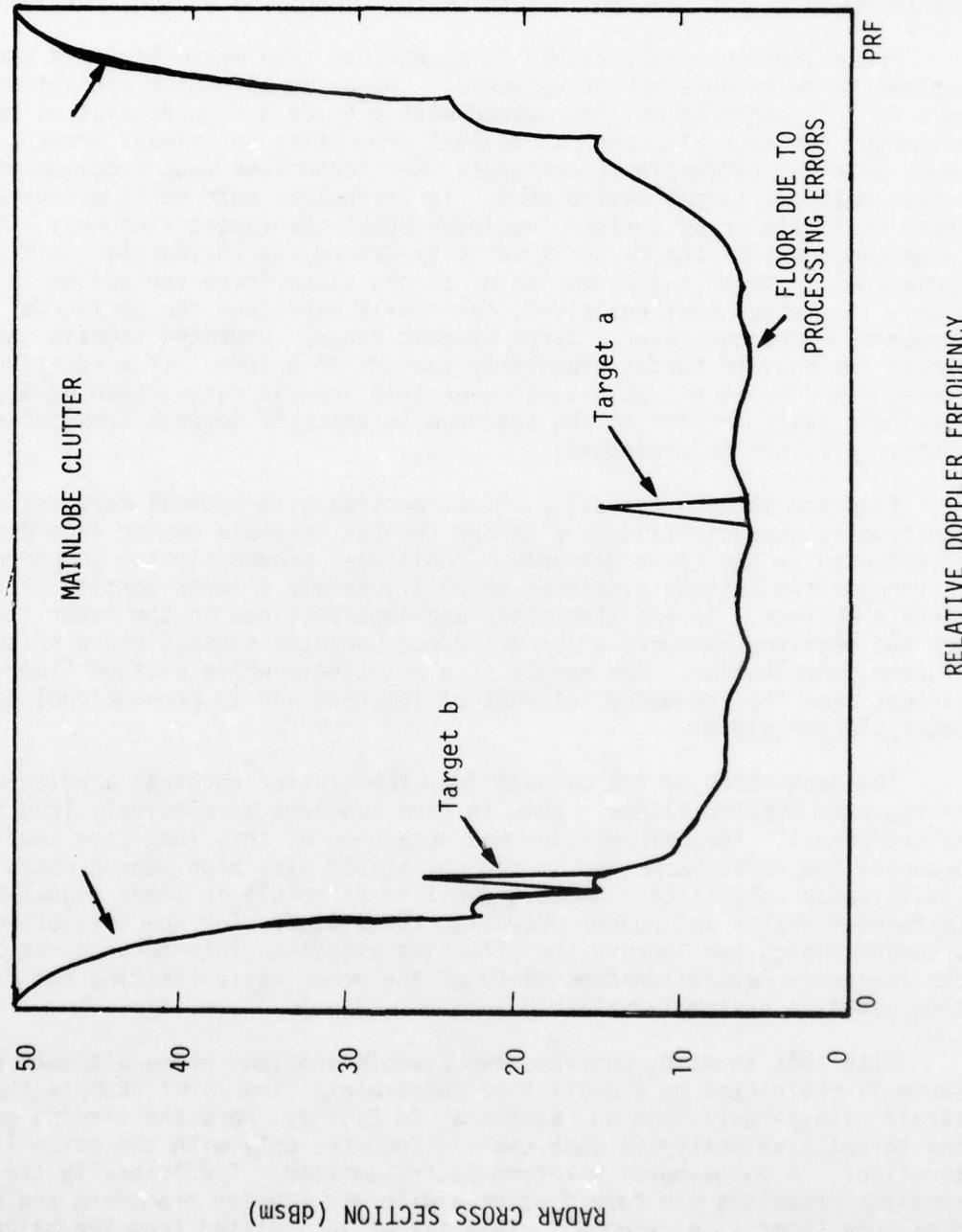


FIGURE 2. Typical Spectrum with Targets and Clutter

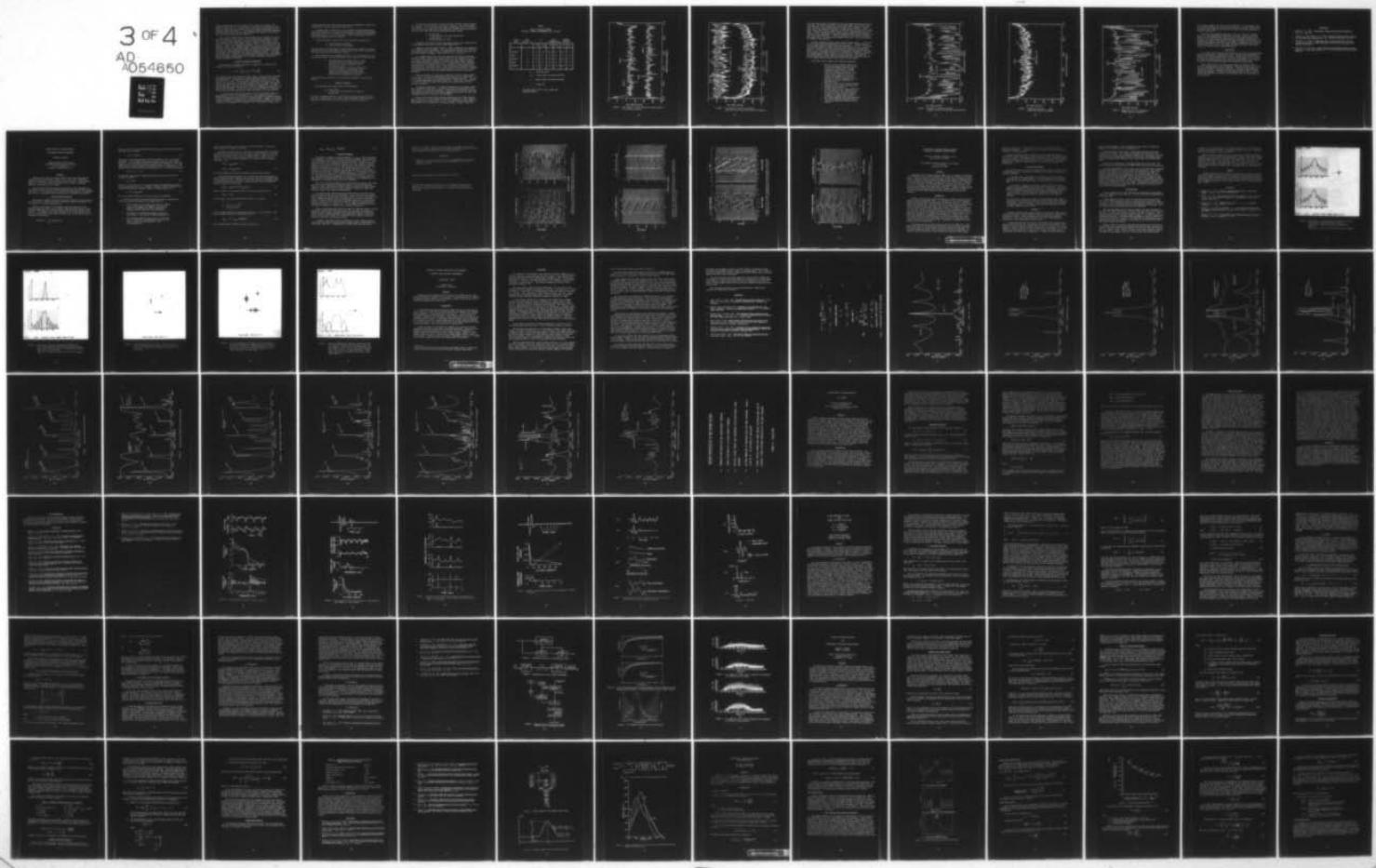
AD-A054 650      ROME AIR DEVELOPMENT CENTER GRIFFISS AFB N Y  
PROCEEDINGS OF THE RADC SPECTRUM ESTIMATION WORKSHOP HELD ON 24--ETC(U)  
MAY 78

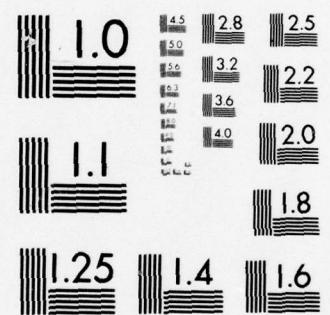
F/G 14/2

UNCLASSIFIED

NL

3 OF 4  
AD-A054650  
SEE PAGE





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

clutter response and the traditional FFT will detect this target. The detection statistics will be determined by the signal-to-interference ratio which is proportional to the dwell time and inversely proportional to the effective filter resolution. If more advanced spectral estimation techniques can narrow the filter then they should increase the signal-to-interference ratio and improve the detection statistics.

This second example concerns those targets such as "target b" in Fig. 2. Here the target is visible against the true spectrum of background clutter but is close enough in Doppler to the main clutter response that a fairly high resolution with large dynamic range is required. The required resolution will require an unacceptably long dwell time. In this example the same waveform is used. This problem stresses the ability to resolve a small signal from a much larger, unwanted signal as contrasted with the first example where the ability to improve the effective signal-to-interference ratio was of interest. Radar performance in these two detection scenario is limited by the effective resolution of the spectral estimation process. The remainder of this paper is devoted to the examination of whether or not performance may be improved by making better use of signal processing, specifically spectral estimation techniques.

#### Spectral Estimation Techniques

By definition, the power spectrum of a signal  $s(t)$  is given by the square of its Fourier transform

$$P(f) = \left[ \int_{-\infty}^{\infty} s(t) e^{-2\pi jft} dt \right]^2 \quad (1)$$

This expression is simple enough and there exists a very efficient algorithm for estimating  $P(f)$  from discrete samples--namely the famous fast Fourier transform or FFT. The problem of course comes about when only a short segment of  $s(t)$  is known. The conventional approach is to assume  $s(t)$  is zero outside the region where it has been sampled or that it repeats periodically and to apply Eqn. (1) using the FFT or perhaps a windowed FFT. In either case the expected value of the result is the true spectrum convolved with the spectrum of the window. The latter gives rise to the familiar diffraction limit, which says that resolution of order less than  $T^{-1}$  are not obtained.

It is well-known that the optimum filter for a bandlimited signal in white gaussian noise is the matched filter. When the desired signal is a Doppler-shifted replica of the original, then the Fourier transform evaluated at the true Doppler frequency acts as a matched filter. However, in practice background clutter is not white and in fact a weighted (tapered) DFT is often superior to the ordinary DFT or matched filter. Our results indicate that

linear prediction and maximum entropy may be good candidates to replace the weighted DFT or perhaps even the matched filter.

In recent years, alternatives to the windowed transform approach have been forwarded which do not require the unrealistic assumption that unknown data equals zero.[1-3] Of these, the most appealing for the above detection problem are those for which no additional a priori information, such as a restricted frequency band, are required.

The two techniques singled out for this feasibility exercise are

1. maximum entropy techniques
2. linear prediction techniques.

Bowling [4] has shown that both of these approaches are superior to the conventional one in (a) resolving two closely space signals and (b) estimating the true frequency of truncated time signals.

Those interested in the mathematical details of the techniques are referred to the work of Bowling and others.[1,2] Suffice it to say here that

- a. The maximum entropy method produces a power spectrum which is consistent with existing data and is maximally noncommittal about missing data, and
- b. The linear extrapolation method assumes that the missing data can be predicted by autoregressive techniques and that the power spectrum can be estimated by Fourier transforming the extrapolated time sequence.

These techniques are applied to the two sample problems in the following section.

#### Numerical Examples

Two representative clutter spectra are considered,

- a. white noise
- b. a large Gaussian spectrum with a floor of white noise.

Case (b) is the more typical of a look-down surveillance scenario, while case (a) is appropriate if the signal is passed through a whitening filter designed to output white noise if the input is similar to case (b).

The procedure followed was to select the desired power spectrum, generate a signal with that spectrum, then add a target of a specified strength and frequency. The signal was then inverse transformed using a 256 point FFT. The first 64 points were used as the known data and the rest were discarded. These 64 points were then spectrum analyzed by three methods:

1. 64-point FFT
2. Linear prediction (50 lags, extrapolated to 256 points)
3. Maximum Entropy (50 lags).

The results consisted of plots of the power spectra and a determination of the signal-to-mean noise and the signal-to-peak noise.

Figures 3-7 present typical results. Table I summarizes the results for five different cases considered. The signal-to-mean clutter ratio ( $S/\bar{C}$ ) and the signal-to-peak clutter ( $S/C_M$ ) are presented for the three spectral estimation techniques considered. Each result is for a single run.

It is appropriate to say a bit more about how the various techniques should be judged. Usually the signal processor can be thought of as a bank of filters. The output of these filters are then used to make decisions about whether targets of interest are present, how many targets, and what are their positions and velocities, etc. In order to do so, one must know the probability distribution of the filter output for targets plus interference and for interference alone. These distributions determine the detection probability for a given false alarm probability or vice versa. Naturally the techniques which yield the higher detection probability are more desirable.

There are two ways of assessing the detection and false alarm probabilities. One is by deriving mathematical expressions for the probability distributions for signal plus interference and interference alone. The other is to perform experiments and gather statistics. These would then be used to infer the underlying detection statistics and overall merits of the various techniques.

The approach taken here is to perform experiments using computer generated targets and interference. Insufficient results are available at this time, however, to infer any statistics. Therefore we shall merely point out some general features which consistently appear in every case run, pose some unanswered questions, and offer a few hypotheses.

Figure 3 shows two original spectra (256 points) with targets and two types of clutter, colored (example 1) and white (example 2). The signal-to-mean clutter level in each case is 16 dB. In calculating the mean clutter, the mainlobe is not considered in the colored cases. Figure 4 shows the

TABLE I  
TABLE OF RESULTS SHOWING  
SPECTRAL ESTIMATION PERFORMANCE FOR FIVE CASES

CLUTTER TAPE	TARGET DOPPLER * $T_S$	FFT	LINEAR EXTRAPOLATION		MAXIMUM ENTROPY		
		S/ $\bar{C}$	S/C <sub>M</sub>	S/ $\bar{C}$	S/C <sub>M</sub>	S/ $\bar{C}$	S/C <sub>M</sub>
COLORED	.4	2.4	.35	26.8	1.5	24.2	1.6
WHITE	.3	6.1	1.3	36.5	2.0	17.1	3.7
COLORED	.3	9.2	1.4	25.6	1.8	44.8	2.5
WHITE	.45	7.3	1.2	21.0	1.6	11.1	1.3
COLORED	.5	4.7	1.3	38.2	3.16	118.0	3.0

$T_S$  = sampling period

$C_M$  = clutter peak (excluding mainlobe)

$\bar{C}$  = mean clutter (excluding mainlobe)

S = target RCS

The above refer to the clutter content per frequency sample.

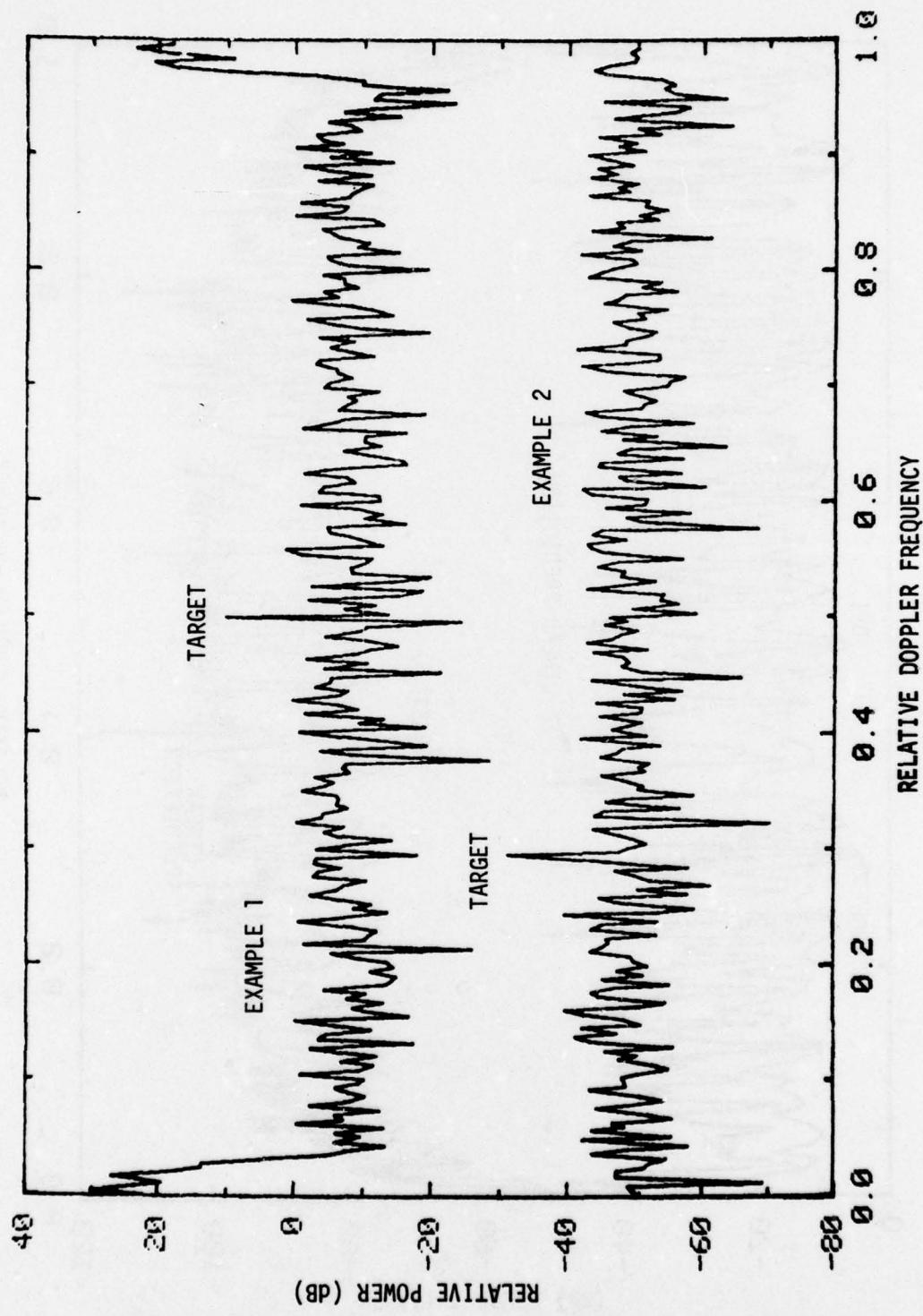


FIGURE 3. Two Examples of True Target and Clutter Spectra:  
(1) Colored; (2) White

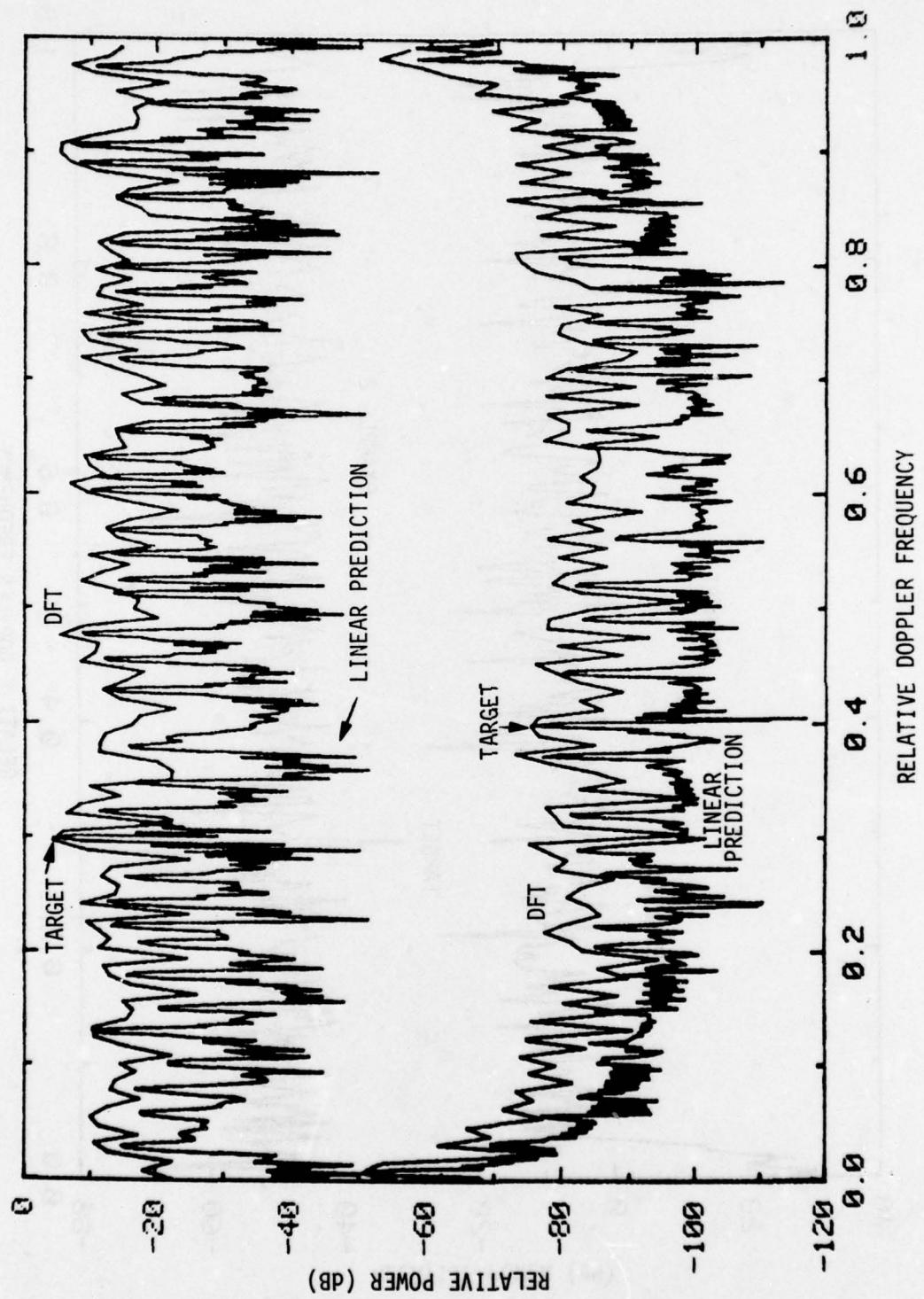


FIGURE 4. Two Examples of Spectral Estimates  
Comparing the DFT and Linear Prediction Methods

resulting spectra when the 64-point DFT and the linear-prediction techniques are used. The scales are adjusted so that the targets are at exactly the same intensity. This permits the clutter peaks to be compared. In both cases the linear-prediction spectrum is entirely below that of the DFT. The mean clutter level is well below that for the DFT and the peaks are slightly lower. Figure 5 compares the DFT spectrum and a maximum entropy spectrum overlaid in the same manner. Again the maximum entropy clutter spectrum lies well below that of the DFT. It would be wrong to infer that this improvement over the DFT will always exist however since this is only one statistical sample of the clutter.

Figure 6 shows the results using a second clutter sample. In this sample the linear prediction method does not do as well. It created several clutter peaks which are higher than those generated in the DFT. This points out that the problem is highly statistical and requires a more extensive analysis. Figure 7 shows the maximum entropy spectrum for the same case shown in Fig. 6 and again several noise frequencies are enhanced. Not shown but included in the summary table is a case which was identical except for the random numbers used to generate the true spectrum. There both the maximum entropy and linear prediction method were more successful than the DFT in suppressing the clutter peaks relative to the target.

The results seem to exhibit the following properties:

1. Both the linear prediction method and the maximum entropy techniques seem to select and accentuate the strongest frequencies which appear in the 64-point DFT spectrum. Except for the target and the mainlobe clutter, if it exists, there appears to be little correlation between the strong frequencies in the true spectrum (Fig. 3) and those of linear prediction or maximum entropy.
2. There is a very high correlation between the peaks of the linear prediction and maximum entropy spectra.
3. In most cases both the linear prediction and the maximum entropy techniques significantly improve the signal-to-mean clutter ratio and provide a modest improvement in signal-to-peak clutter ratio when compared with the 64-point DFT
4. The clutter spectrum shape does not provide any noticeable difference in the results.

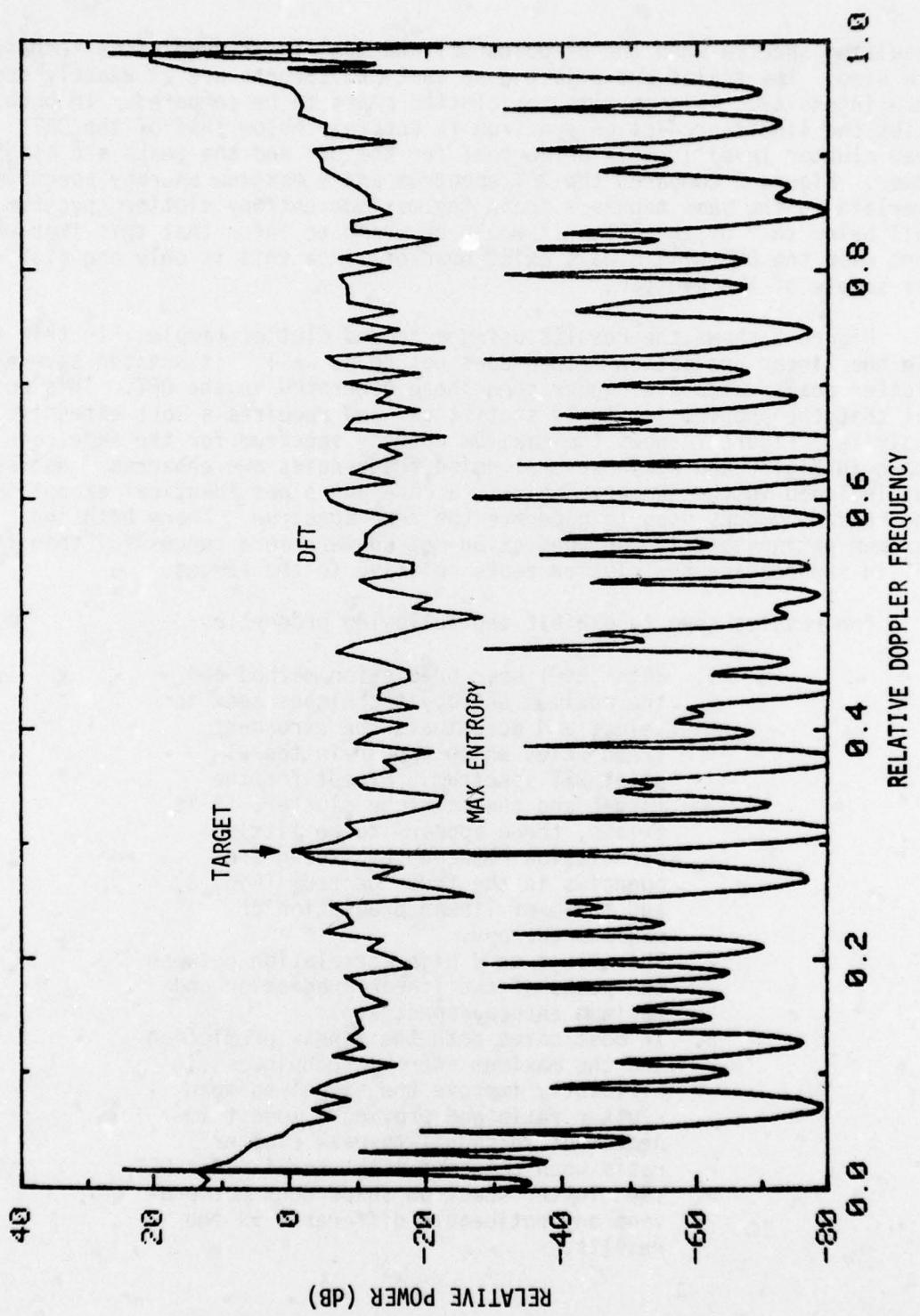


FIGURE 5. Example of A Spectral Estimate  
Comparing the DFT and Maximum Entropy Methods

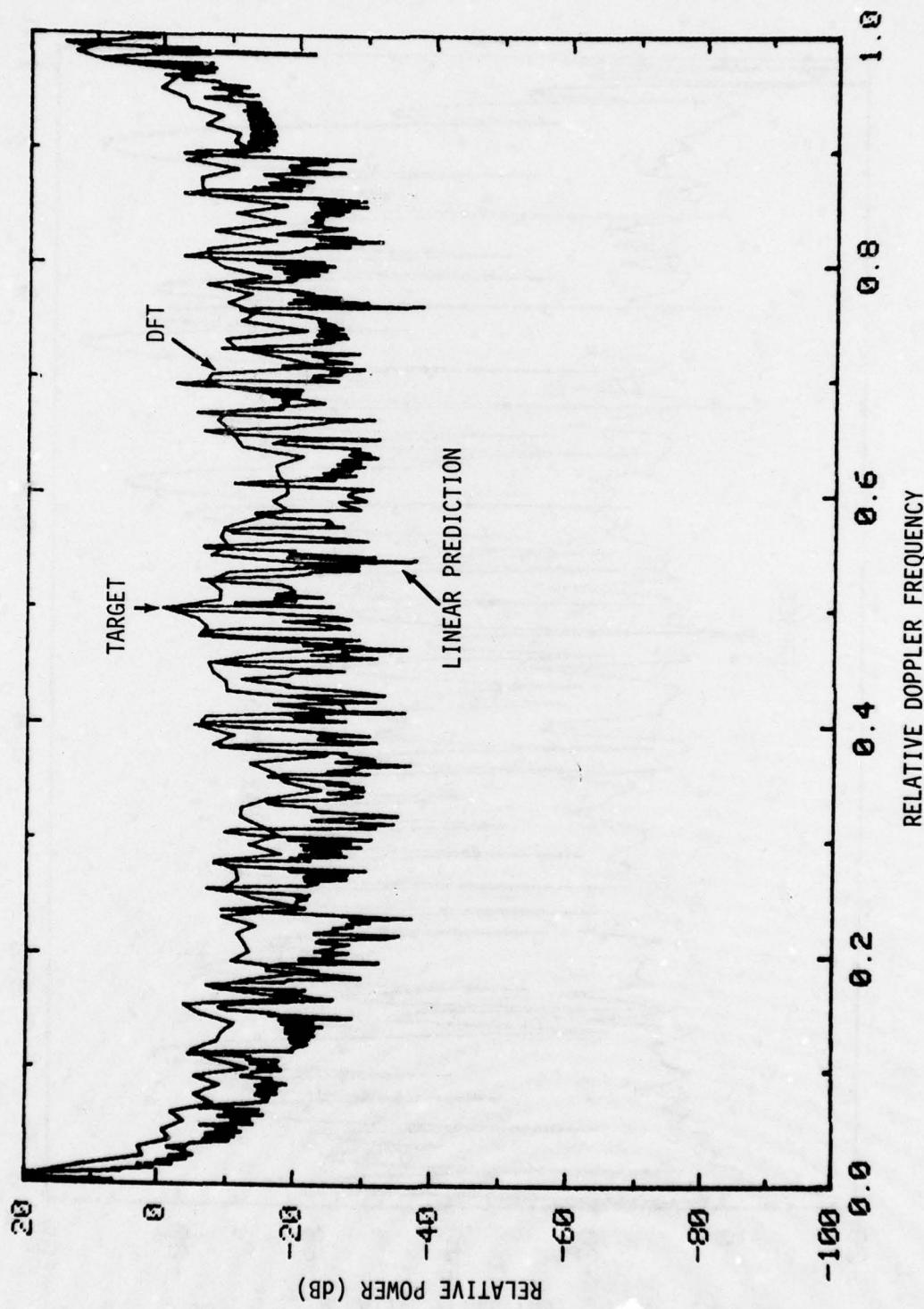


FIGURE 6. Comparison of A DFT and Linear Prediction Spectral Estimation

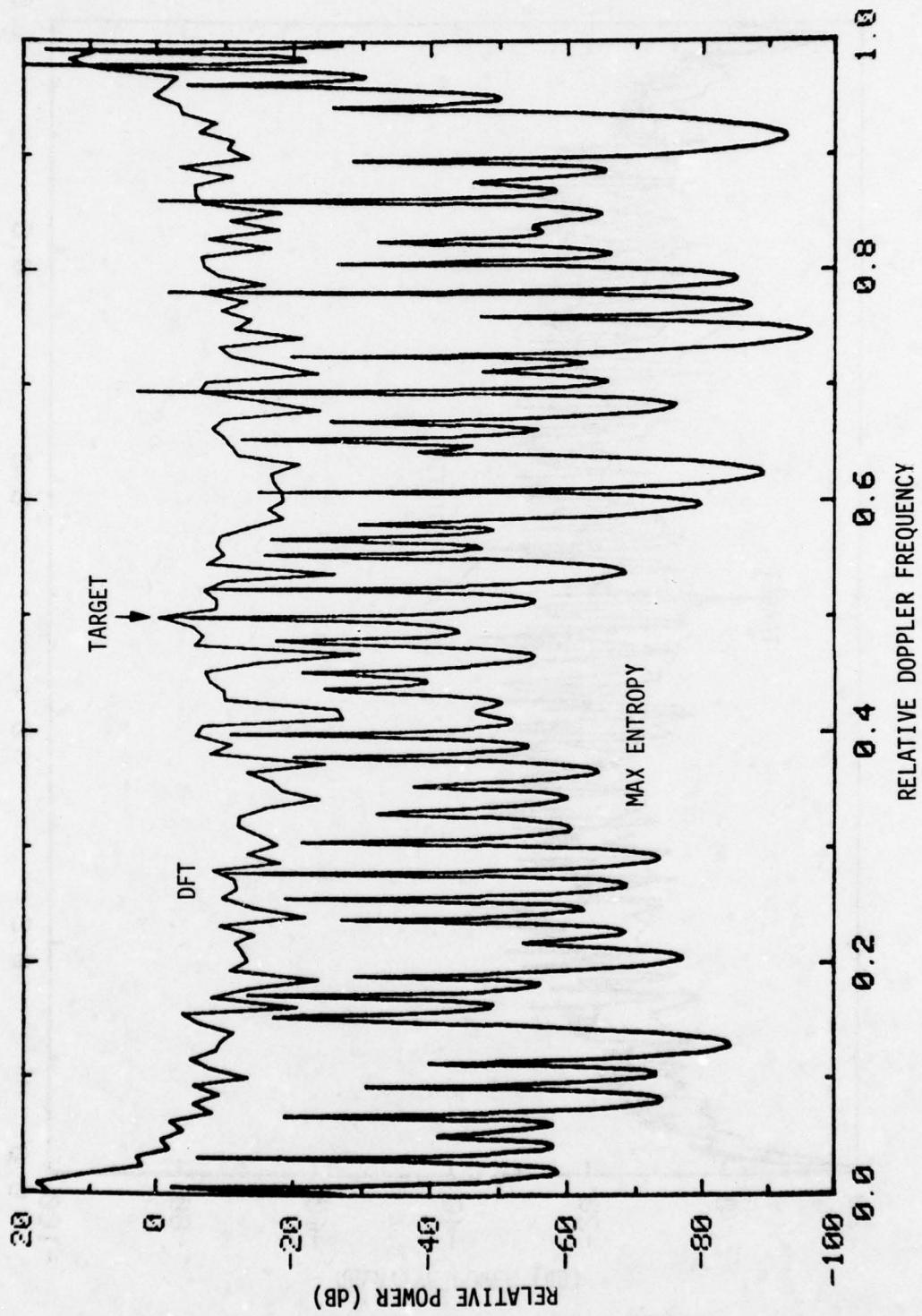


FIGURE 7. Comparison of A DFT and Maximum Entropy Spectral Estimation

Of the above comments, the third is most noteworthy. If the advanced spectral estimation techniques can provide an effective increase in the signal-to-clutter ratio, then they would be useful signal processing approaches in detection applications.

There are some unanswered questions. The first question is whether there is a real improvement in detection statistics. Statistical distributions for the clutter after processing will be required to answer this question. Are there better spectral estimation techniques than the one considered here? What is the impact of the number of expansion coefficients on detection performance? How far can the signal be expanded and still get improved performance? What is the impact of signal-to-clutter ratio before processing on the utility of these techniques?

#### Conclusions

As is typical of most preliminary exercises, this one has generated more questions than answers. However, the results indicate some improvement is possible and should encourage further study. Taking these results at face value one can conclude that there is a possible improvement of at least one dB signal to clutter (based on the signal-to-maximum clutter ratio) using either the linear prediction method or the maximum entropy method.

These results are all the more positive when one considers that in practice a tapered window such as Hamming or Chebyshev would be used in conjunction with the DFT to suppress sidelobes. A tapered window always decreases the effective resolution and the effective dwell time for clutter suppression purposes. Therefore any method which does as well as the un-tapered DFT and does not produce a high sidelobe level will be superior to the tapered DFT.

### References

1. Jaynes, E.T., (1957), "Information Theory and Statistical Mechanics," Phys. Rev. 106, 620.
2. Ulrych, T.J. and Bishop, T.N., (1975), "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geophys. Space Phys. 13, 183.
3. Papoulis, A., (1975), "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," IEEE Transactions on Circuits and Systems, Vol. CAS-22, No. 9, 735.
4. Bowling, S.B., May 1977, "Linear Prediction and Maximum Entropy Spectral Analysis for Radar Applications," M.I.T. Lincoln Laboratory Project Report RMP-122.

RADAR IMAGING OF DISCRETE TARGETS  
WITH MAXIMUM ENTROPY TECHNIQUES\*

STEPHEN B. BOWLING

Radar Signature Studies Group  
Massachusetts Institute of Technology  
Lincoln Laboratory  
Lexington, Massachusetts 02173

Abstract

Maximum entropy spectral analysis (MESA) offers some advantages for imaging discrete targets when a data set is too short for conventional Fourier procedures to resolve important spectral features. Radar measurements can, for example, provide estimates of the length, cross-range dimension, and spin rate of an object.

This paper first derives the specifications for radar waveforms with which target features can be estimated satisfactorily using standard Fourier techniques. If the waveform is not optimal, then MESA may help to compensate for the loss in resolution otherwise sustained in conventional processing.

1. Radar Imaging Principles

Imaging with standard techniques of spectral analysis has been discussed in the literature, although the use of alternative spectral techniques for this purpose is a recent application [1].

Sizing a radar target in cross-range relies upon detecting those frequency components present in a coherent burst of pulses that are correlated with the motion of scattering centers relative to the target's center of motion. For example, if a range resolution cell contains  $N$  scattering centers, then the total radar return from that cell is the coherent sum of  $N$  individual contributions

$$A \exp[i\phi(t)] = \sum_{n=1}^N A_n \exp[i\phi_n(t)] \quad (1)$$

where  $A_n$  and  $\phi_n(t)$  are the amplitudes and phases of the  $N$  scattering centers. The phase  $\phi_n(t)$  is simply

$$\phi_n(t) = 4\pi R_n(t)/\lambda \quad (2)$$

where  $R_n(t)$  is the one-way range to the  $n$ th scatterer, and  $\lambda$  is the radar wavelength. If the scatterer is spinning about an axis, then the range changes in time to produce a variation in phase that appears as a sinusoidal component in a series of consecutive samples (pulses) of Equation (1). (We have assumed any accelerations are negligible during the observation period, although such could be detected and accounted for.) If we write the range as

$$R_n(t) = v_n t + R_{no} \quad (3)$$

we find that the frequency component produced by the motion of the  $n$ th scatterer is simply

$$f_n = 2v_n/\lambda \quad (4)$$

where  $v_n$  is the velocity of the scatterer during the observation interval. If the angular spin rate  $\omega_{spin}$  (rad/sec) is known, then the distance of the scatterer from the spin axis, as viewed at an aspect angle  $\Omega$ , is

$$r_n = v_n / (\omega_{spin} \sin \Omega) \quad (5)$$

and a cross-range dimension in a metric unit can be estimated.

Several basic signal processing considerations must be satisfied in order to obtain good estimates of the moment arms  $r_n$ :

- a. the data must not be aliased, such that the sampling rate or pulse repetition frequency (PRF) must be chosen to accommodate the highest frequency (or maximum velocity) expected to be encountered;
- b. the length of the observation period  $T_p$  must be long enough to resolve each frequency component  $f_n$ ;
- c. the observation period  $T_p$  must not be so long that non-linearities (accelerations) cause the frequency spectrum to smear because of time non-stationarities.

These three requirements will determine the radar waveform, initially designed according to Fourier principles.

The Nyquist interval spans  $\pm 1/2\delta t$  (minus to plus the Nyquist frequency) and converts by Equation (4) into a velocity interval spanning  $\pm \lambda/4\delta t$  (the ambiguous velocity interval);  $\delta t$  is the sample spacing. Therefore, the magnitude of the largest velocity must not exceed  $\lambda/4\delta t$  in order to avoid aliasing. If  $L$  is the largest dimension of the target, then we calculate, since  $PRF = 1/\delta t$ ,

$$\lambda/4\delta t = \omega_{\text{spin}} L \sin \Omega, \quad \text{or}$$

$$PRF_{\min} = 4\omega_{\text{spin}} L \sin \Omega / \lambda \quad (6)$$

to be the minimum pulse repetition frequency that prevents aliasing.

During the processing interval  $T_p$ , no scattering center may change velocity resolution cells; otherwise, the spectral lines will be smeared across several resolution cells. Relative to the center of mass, the largest possible change in velocity occurs as the scattering center spins through zero velocity along the line of sight:

$$\Delta V_{\max} = 2\omega_{\text{spin}} L \sin \Omega \sin(\omega_{\text{spin}} T_p / 2) \quad (7)$$

which we require to be confined within the velocity resolution cell

$$\delta V_{\text{res}} = \lambda / 2T_p \quad (8)$$

Setting  $\Delta V_{\max} = \delta V_{\text{res}}$  and approximating  $\sin(x) \approx x$ , we obtain

$$T_p = \left[ \frac{\lambda}{2\omega_{\text{spin}}^2 L \sin \Omega} \right]^{1/2} \quad (9)$$

as the optimum length of the observation period (i.e., burst length). This gives a velocity resolution from Equation (8) of

$$\delta V_{\text{res}} = \sqrt{\lambda} (\omega_{\text{spin}} \sqrt{(L \sin \Omega) / 2}) \quad (10)$$

and a minimum number of samples (pulses) during  $T_p$  of

$$N_{\min} = (\text{PRF}_{\min}) T_p = \sqrt{8L \sin \Omega / \lambda} \quad (11)$$

## 2. Practical Examples

Approximate knowledge of target parameters ( $L$ ,  $\omega_{\text{spin}}$ ) and of the observation geometry ( $\Omega$ ) helps in designing the radar waveform. The minimum PRF is calculated by Equation (6), and the temporal length of the burst is calculated by Equation (9). Examples are easily simulated. Imagine a spinning sphere with six scattering centers symmetrically spaced around the equator. Observed at an angle of  $\Omega = 30$  deg from the pole, the spinning scattering centers individually contribute to the Doppler spectrum of the target. At any time, at least three scatterers are visible and hence at least three frequencies are present. Figure 1 demonstrates (a) the aliasing produced by too low a PRF, (b) the lack of resolution because  $T_p$  is too short, (c) correct choices of both PRF and  $T_p$ , and (d) the smearing caused by too long a  $T_p$ . The spinning motion is apparent in Figure 1-C to the extent that the spin period can be estimated by timing the appearance and disappearance of a scatterer as it becomes visible and is then shadowed.

Even if the waveform parameters are chosen properly, Equation (10) shows that the velocity resolution ultimately depends on the radar wavelength. Using too large a wavelength can thwart the most carefully designed imaging waveform. An estimate of the velocity resolution necessary to image the target is useful for selecting the wavelength.

The radar waveform specifications described above will provide satisfactory resolution upon Fourier processing. In the event that the waveform is not optimal for Fourier processing (burst length too short; wavelength too long), maximum entropy may help to compensate for the loss of resolution. Figure 2 compares conventional spectra and MESA spectra when the processing interval is too short to resolve the frequency components produced by the spinning scatterers (cf., Figure 1-B). It is evident that MESA outperforms the Fourier transform in resolving the individual frequencies.

Figure 3 shows field data taken by a radar on a spinning object. In this case, both the wavelength and burst length were too short for Fourier procedures to resolve fully the spectral lines that MESA begins to perceive. Indeed, there are common features in the simulations (Figures 1-C and 2-B) and in the field data (Figure 3-B) indicative of the motion of scattering centers as they spin into and out of view.

In summary, MESA offers definite advantages in radar imaging applications when Fourier processing of a burst waveform fails to provide sufficient cross-range resolution. A combination of linear predictive and Fourier

techniques, in which a data set is extended by a prediction filter before Fourier transformation rather than being padded with zeroes, offers an acceptable compromise that combines the best features of both [1].

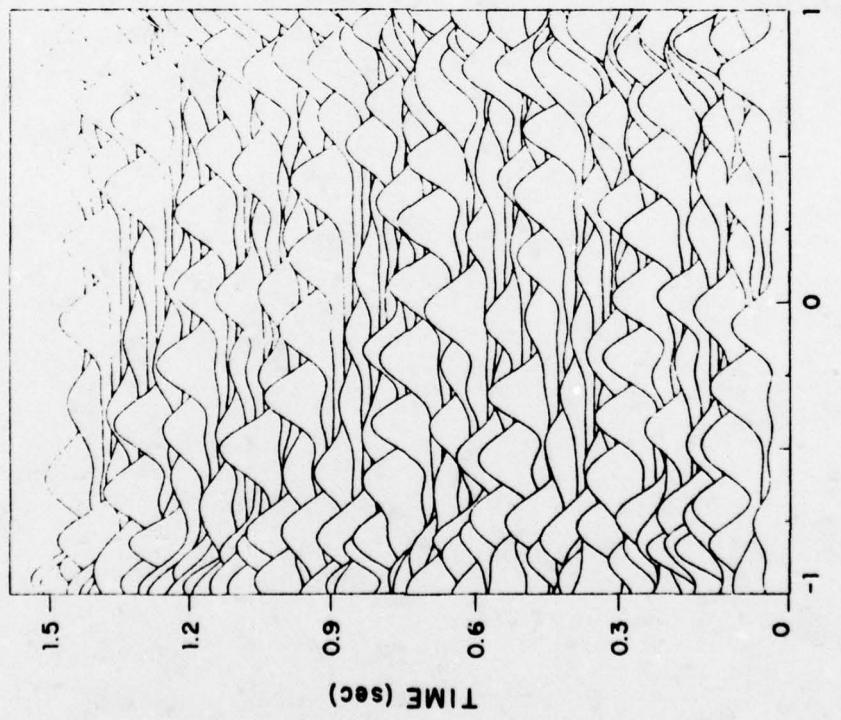
References

1. Bowling, S. B., 1977, "Linear Prediction and Maximum Entropy Spectral Analysis for Radar Applications," Project Report RMP-122, Lincoln Laboratory, M.I.T.

\*This work was sponsored by the Department of the Army.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

(A) ALIASED SAMPLING



(B) DATA SET TOO SHORT

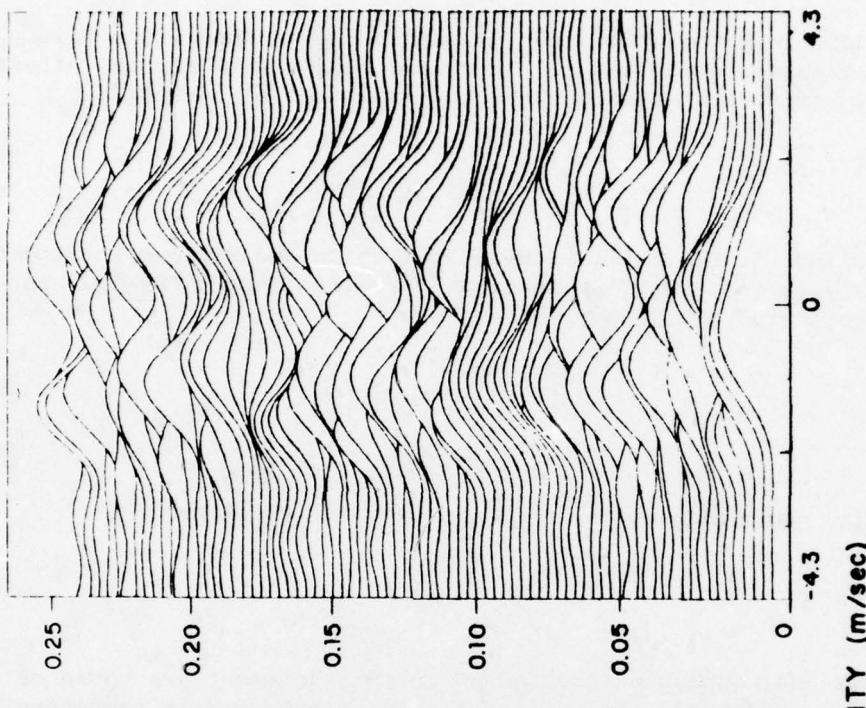
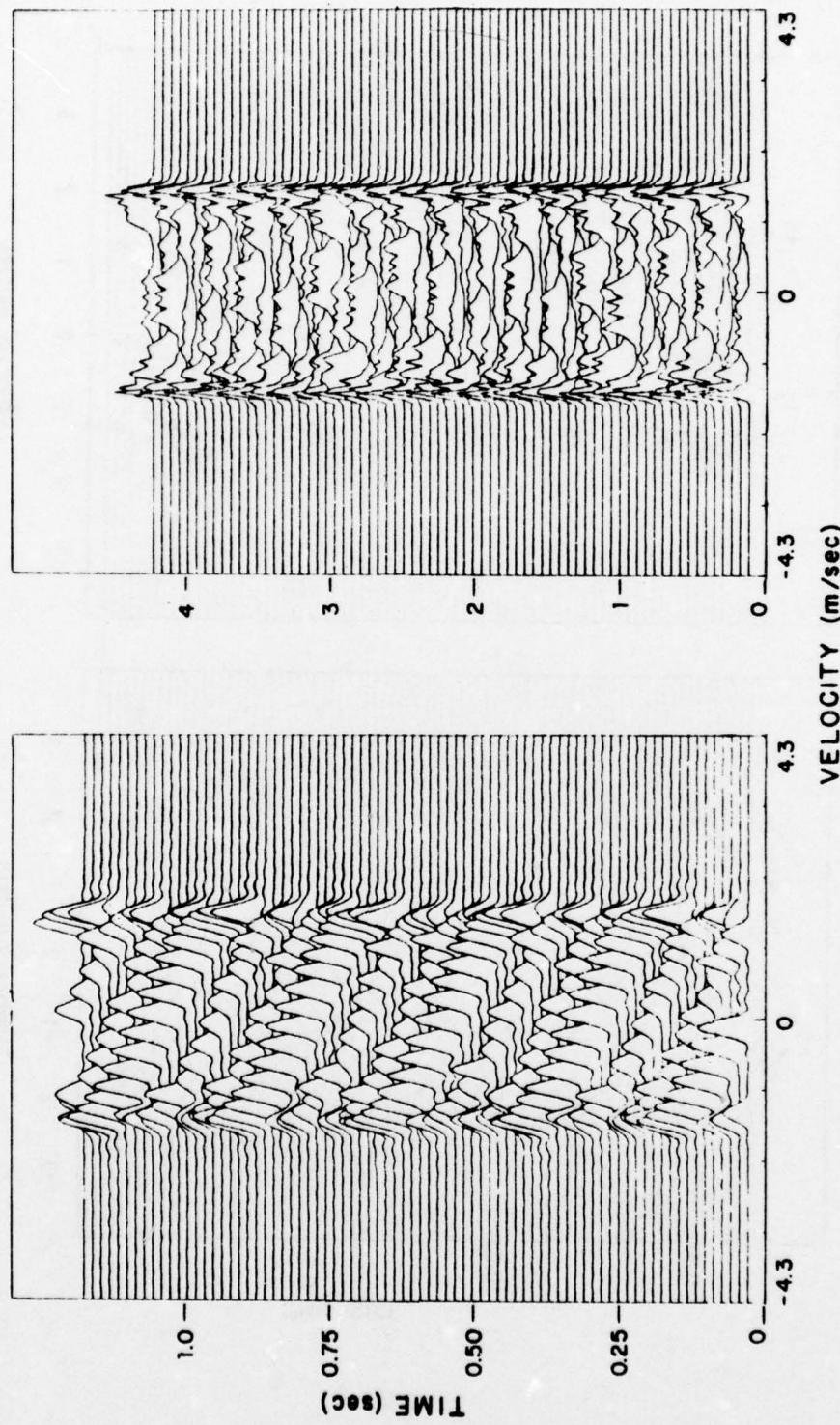


FIGURE 1(a, b). Simulated Doppler signature of a spinning sphere when (a) the PRF is too low and (b) the processing interval is too short.

(C) CORRECT PARAMETERS



(D) DATA SET TOO LONG

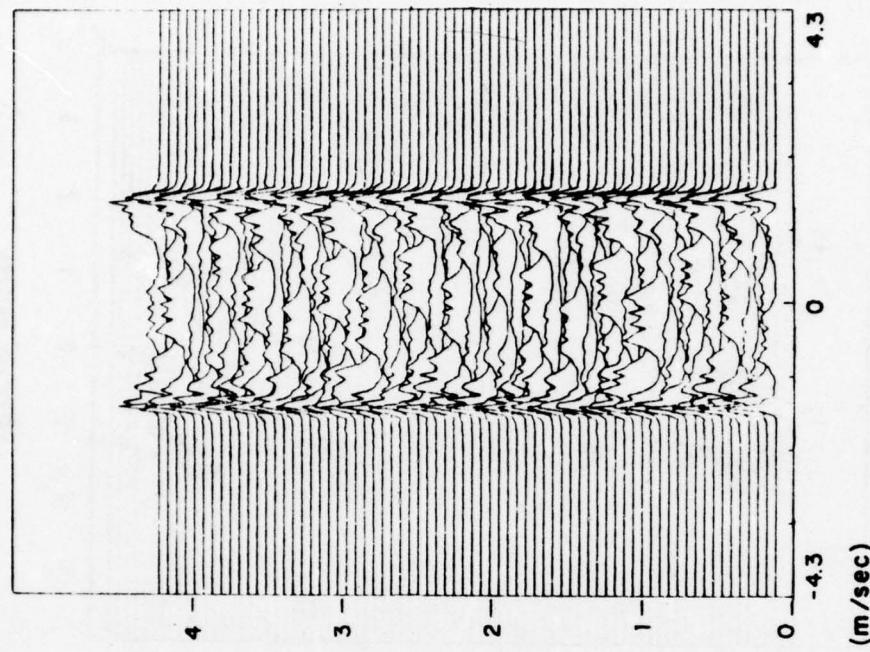


FIGURE 1(c, d). Simulated Doppler signatures of a spinning sphere (c) for the correct choice of waveform parameters and (d) when the processing interval is too long.

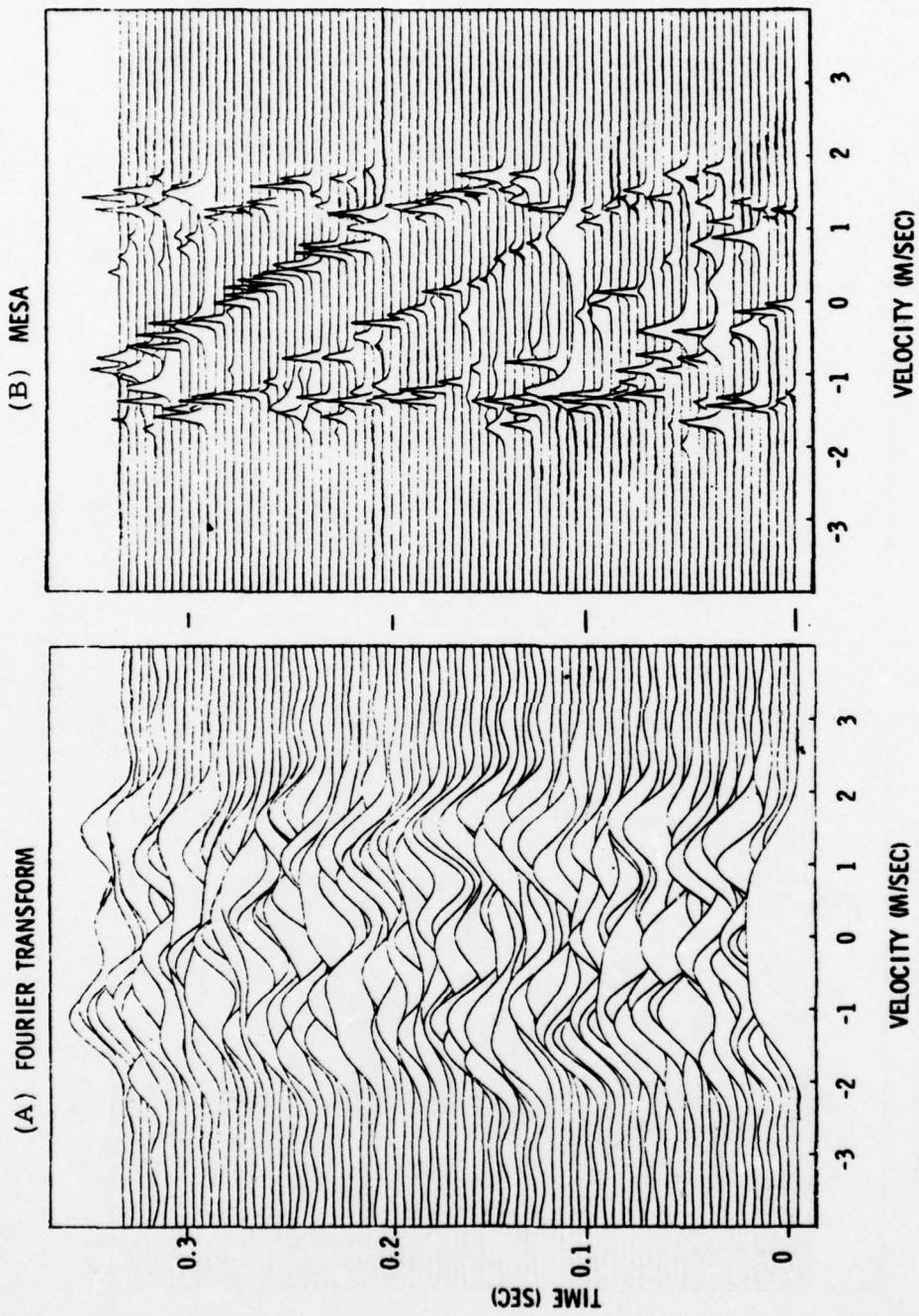


FIGURE 2(a, b). Fourier techniques in (a) do not fully resolve the scattering centers on a spinning sphere which MESA (b) resolves.

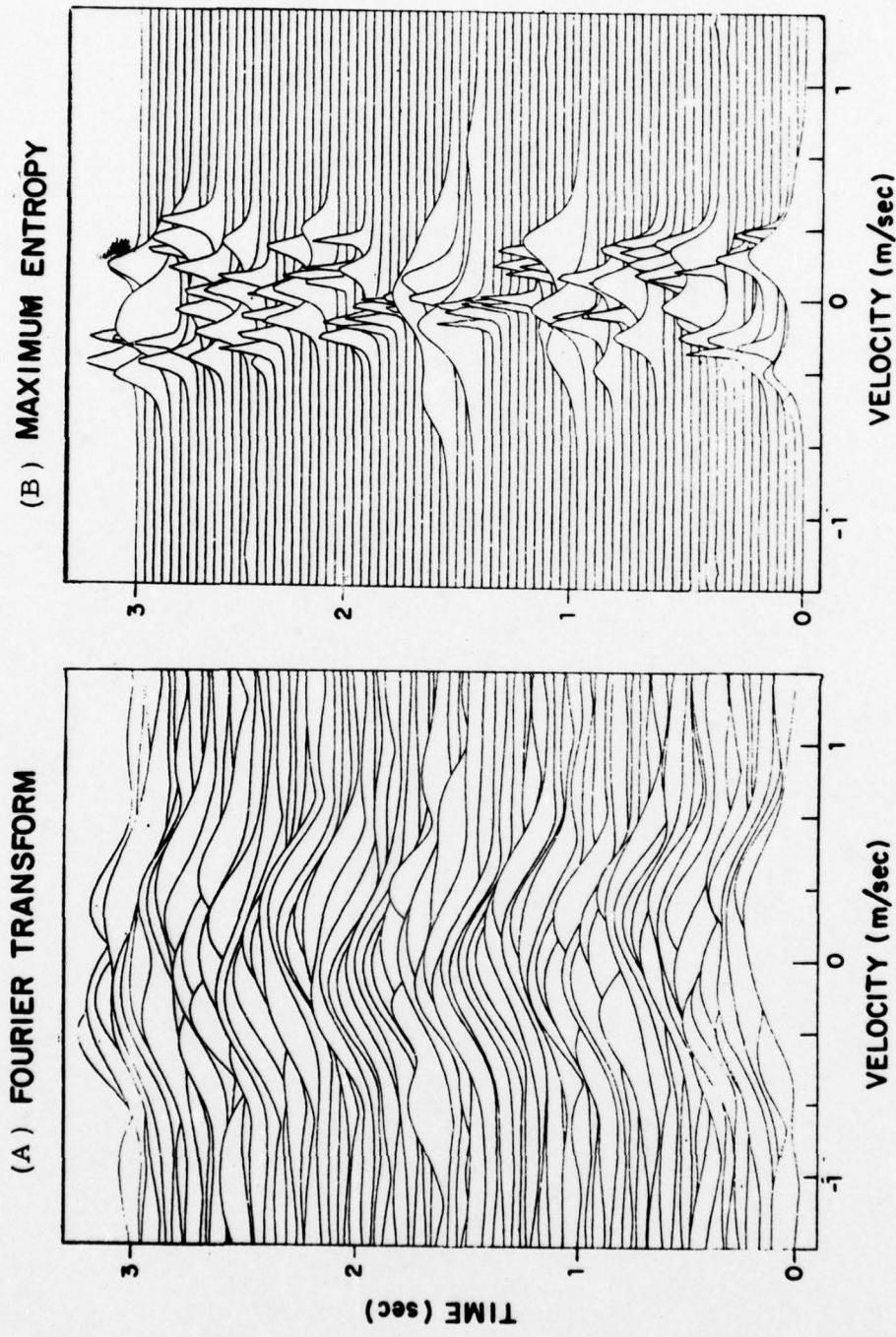


FIGURE 3(a, b). Radar data on a spinning object. MESA begins to resolve the tracks of scattering centers not resolved by Fourier techniques.

APPLICATION OF MAXIMUM ENTROPY FREQUENCY  
ANALYSIS TO SYNTHETIC APERTURE RADAR

Philip L. Jackson, Lawrence S. Joyce,  
and Gerald B. Feldkamp

Environmental Research Institute of Michigan  
P. O. Box 8618  
Ann Arbor, MI 48107

Abstract

Maximum entropy frequency analysis (MEM), one of a number of autoregressive techniques, can be used to improve the resolution of synthetic aperture radar. Limited by Fourier transform theory, current SAR processing procedures produce relatively large mainlobes and significant sidelobes. MEM produces narrow mainlobes and negligible sidelobes with very limited data. Therefore we have the potential of achieving improved resolution with presently used data sets, attaining current resolution with reduced amounts of data, and detecting weak signals now buried under sidelobes. Problems of applying MEM to SAR include one-dimensionality, longer computing time than the FFT, spectral splitting, small deviations from correct amplitude, and non-linearity.

Background

Synthetic Aperture Radar (SAR) processing is currently limited in resolution by Fourier transform theory [1]. Recording of the linear frequency modulated range and Doppler pulses requires frequency analysis to obtain an image from the SAR data. The half-power width of the Fourier transform mainlobe determines Rayleigh resolution, and the sidelobes produced by the Fourier transform can obscure nearby weak targets. SAR images are produced with the data limited by the bandwidths of the chirp and Doppler signatures. As a reduction in data produces an increase in the mainlobe of the Fourier transform, resolution is consequently degraded with a reduced amount of data. Hence, SAR resolution is dependent on the amount of available data. Any alternative method of processing which can increase resolution or achieve the current resolution with fewer data would increase the usefulness and possibly extend the applications of SAR. Reduction or elimination of sidelobes would also be beneficial.

The maximum entropy frequency analysis method [2,3] (MEM, also called MESA) can be used to produce narrower mainlobes than the Fourier transform on limited data sets, in addition to virtually eliminating sidelobes. We have applied MEM algorithms to SAR data and shown that, in comparison to Fourier transform processing, finer resolution is obtained with the same

data, this resolution is retained with fewer data, and sidelobes were virtually eliminated . MEM is related to a number of recently developed methods of frequency analysis.

Synthetic aperture radar is primarily used for producing images of terrain and cultural objects which appear much like aerial photography taken at a low sun angle. SAR images are interpreted visually, so that the tones, textures and patterns are the predominant information desired. The precise magnitude of a return is seldom measured, as the current SAR systems are uncalibrated.

#### Potential Improvement

Each of the three areas for which MEM processing offers potential would provide improvement in the amount of information or enable more extensive or economical systems:

The first area - improvement in resolution - would aid in the identification of smaller objects. For one dimension, our experiments have indicated an improvement factor of approximately six when using MEM.

The second - retaining the current resolution with less data - would aid in the gathering of information from satellites. Data transmission requirements are currently excessive. Indications are that 1/6 to 1/7 of data for conventionally processing one-dimensional data is required by MEM. If a two-dimensional algorithm is developed, the amount of data transmission might be reduced to 1/49 of that now required.

The third - detection of weakly reflecting objects near strong reflectors - would aid in detecting many objects which are now obscured by sidelobes. Our experiments on radar data have shown that sidelobes virtually disappear in the image, both from point reflectors and from extended objects with clutter [4].

#### Problems

Although experiments have shown that MEM can improve SAR image process- in these three areas, several problems require answers before application to operational systems. These problems are spectral splitting, moderate deviation (< 3 dB) from correct amplitude, the lack of two-dimensional algorithms, and the computational time required.

At times MEM processing will produce two closely spaced spectral peaks where the data has only one. We have not observed this "splitting" on the SAR data we have processed, probably because the split peaks are very close together. However, splitting could be deleterious to the image. If it were known under what conditions, or what objects produce splitting, then the interpretive procedure could accommodate it. Indeed, the splitting itself

might become diagnostic. Also, development on the theory and algorithm appears to have eliminated this disadvantage [4].

Moderate deviations from correct amplitudes in spectral peaks has a minimal effect on SAR. SAR produces a photograph-like image for visual inspection. As the eye is a logarithmic detector and is primarily concerned with differences in brightness, texture, and patterns, the moderate deviations in the amplitudes of MEM peaks will seldom cause visual misinterpretation.

Noise is more destructive to MEM computations than to the FFT. The effect of noise on operational use of MEM for SAR will require investigation.

A serious disadvantage is the current lack of two-dimensional algorithms, although a development of the theory has been presented [5]. Because MEM does not retain phase, only one direction can be processed by MEM. The method we have used is to first process in one direction with the fast Fourier transform (FFT) and subsequently to process the FFT result in the perpendicular direction with MEM. In this way either the azimuth or range directions can have the benefit of MEM processing. A cumbersome method is to process the original data twice, so that the azimuth data is enhanced in one image and the range data in the other image. Superimposition with point-by-point multiplication and normalization will then enhance the image in both the azimuth and range directions.

#### Illustrations

In this presentation we only illustrate the processing of data synthesized to be similar to actual SAR data. "Point" sources are synthesized and processed by MEM.

A SAR-synthesized "point" source was processed in both directions by FFT. The intensity modulated result on a CRT is shown in the upper right of Figure 1. The horizontal and vertical sidelobes are evident. In the left portion of Figure 1, horizontal and vertical intensities are graphed. Note the width of the mainlobe and the size of the sidelobes.

The same synthesized data as used for Figure 1 was then processed by the FFT in the vertical direction, and by the MEM algorithm in the horizontal direction. A similar CRT display to Figure 1 is shown in Figure 2. The vertical graph is similar to those shown in Figure 1, with wide mainlobe and significant sidelobes. The MEM-processed horizontal direction is shown in the upper left portion of Figure 2. The sidelobes have disappeared, and the mainlobe has become much narrower, particularly at and near the half-power width which governs resolution.

Figure 3 shows the intensity-modulated CRT presentation of an array of "point" reflectors precessed from synthetic SAR data of  $128 \times 128$  points by the FFT. The two closely spaced reflectors in the lower right are resolved with this number of data. Figure 4 shows the same array processed from  $1/4$  the number of data samples ( $64 \times 64$ ). Note that the two closely spaced

reflectors in the lower right are no longer resolved with an FFT processing of this reduced amount of data. The sample spacing is the same. Contiguous samples were removed to reduce the data set.

When this reduced amount of data is processed by the MEM algorithm, the resolution of the two reflectors is maintained. Figure 5 shows the intensity modulated display of the 64 x 64 data points processed by FFT in the vertical direction and by MEM in the horizontal. The upper graph in Figure 5 is a trace across the three lower right reflectors. The two closely spaced reflectors are clearly resolved by MEM. The lower graph is a vertical trace across a main FFT lobe to display the size produced with the data reduced to 64 samples.

#### Summary

The MEM algorithm has been shown to be applicable to synthetic aperture radar data. Better resolution than that found with the Fourier transform is attained with MEM, and resolution is retained with a reduced amount of data. Also sidelobes are virtually eliminated. Problems include spectral splitting, minor deviations from true values, computing time, noise, and one-dimensionality of MEM.

#### References

1. Harger, R. O., 1970, "Synthetic Aperture Radar System: Theory and Design", Academic Press, Inc., New York, NY.
2. Burg, J. P., 1967, "Maximum Entropy Spectral Analysis", Paper presented to 37th Annual Meeting of the Society of Exploration Geophysicists, Oklahoma City, OK.
3. Ulrych, T. J., 1972, "Maximum Entropy Power Spectrum of Truncated Sinusoids", J. Geophys. Res., 77, 1396.
4. Fougere, P. F., 1977, "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", J. Geophys. Res., 82, 1051.
5. Newman, W. I., 1977, "A New Method of Multidimensional Power Spectral Analysis", Astron. Astrophys., 54, 369.

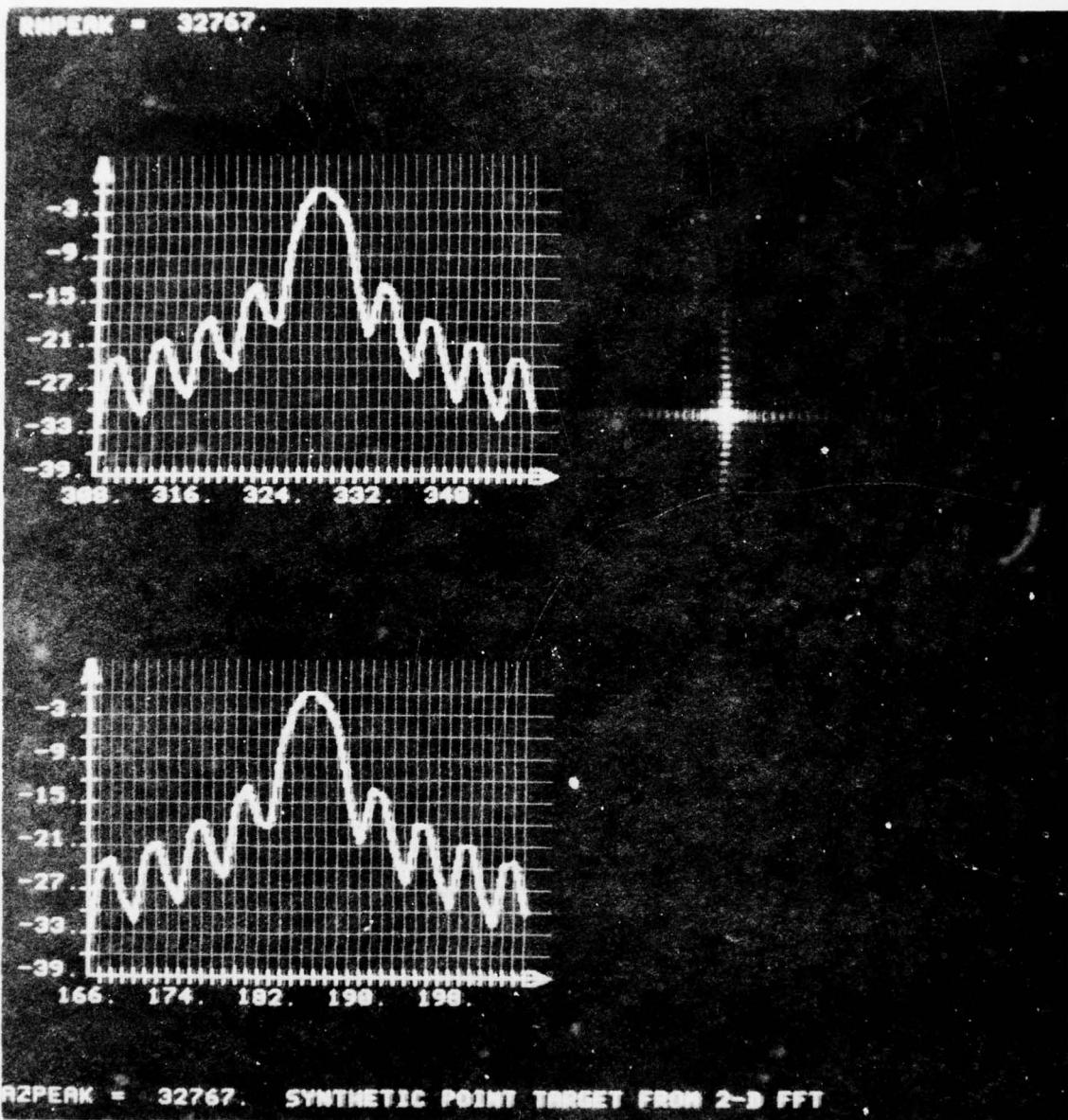
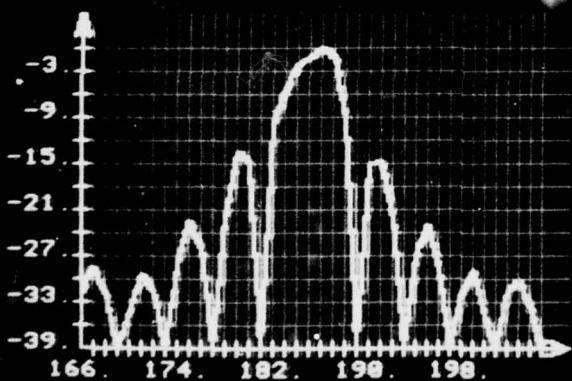
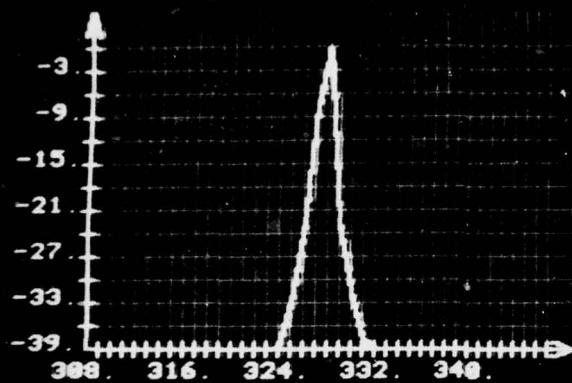


FIGURE 1. Upper Right: Intensity Modulated CRT Display of FFT Processing of a Synthetic "Point" Target.  
Upper Left: Graph of Intensity Along Horizontal Segment.  
Lower Left: Graph of Intensity Along Vertical Segment.

RNPEAK = 32000.



AZPEAK = 32000. SYNTHETIC POINT TARGET FROM FFT/MEM

FIGURE 2. Upper Right: Intensity Modulated CRT Display of Horizontal MEM Processing and Vertical FFT Processing.  
Upper Left: Graph of Intensity Along Horizontal Segment (Note absence of sidelobes).  
Lower Left: Graph of Intensity Along Vertical Segment.

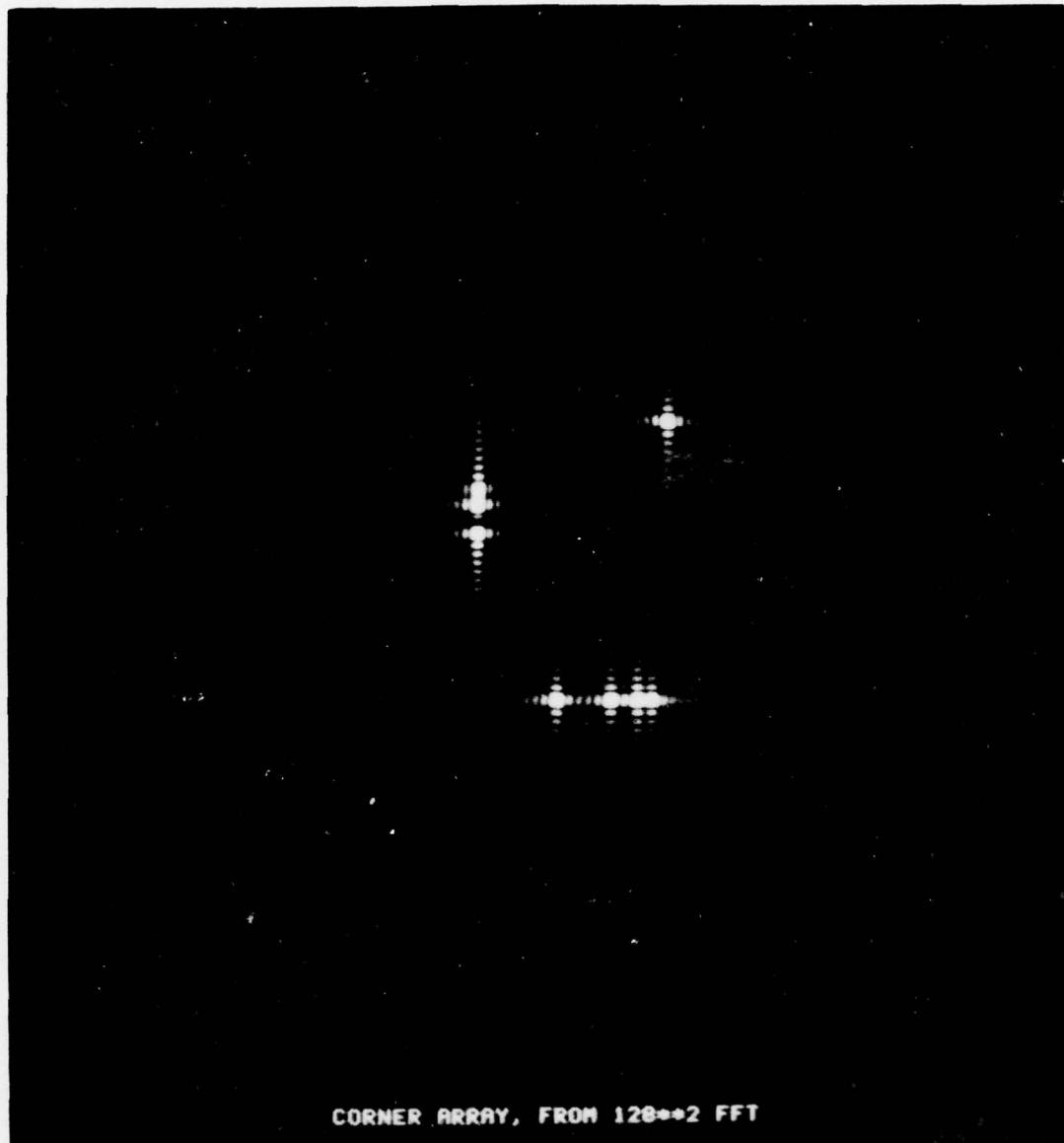


FIGURE 3. Intensity Modulated CRT Display. FFT Processing of 128 x 128 Data Samples of Synthesized Array of "Point" Reflectors. Note Closely Spaced Pair of Reflectors in Lower Right.

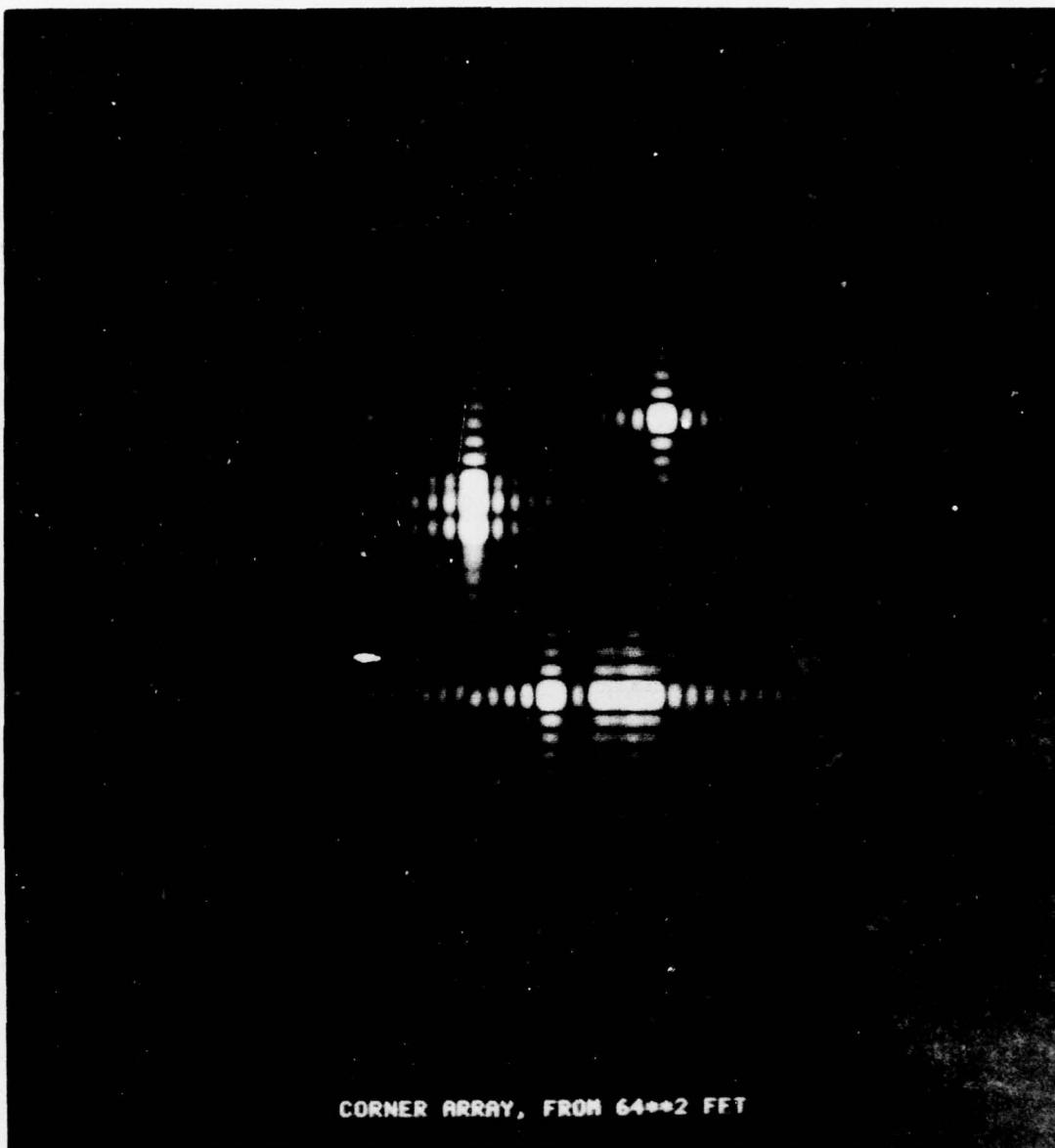


FIGURE 4. Intensity Modulated CRT Display of FFT Processing of 64 x 64 Data Samples of Synthesized Array of "Point" Reflectors. Contiguous Samples Removed from 128 x 128 Data Array Used for Result Shown in Figure 3. Note Loss of Resolution in Lower Right.

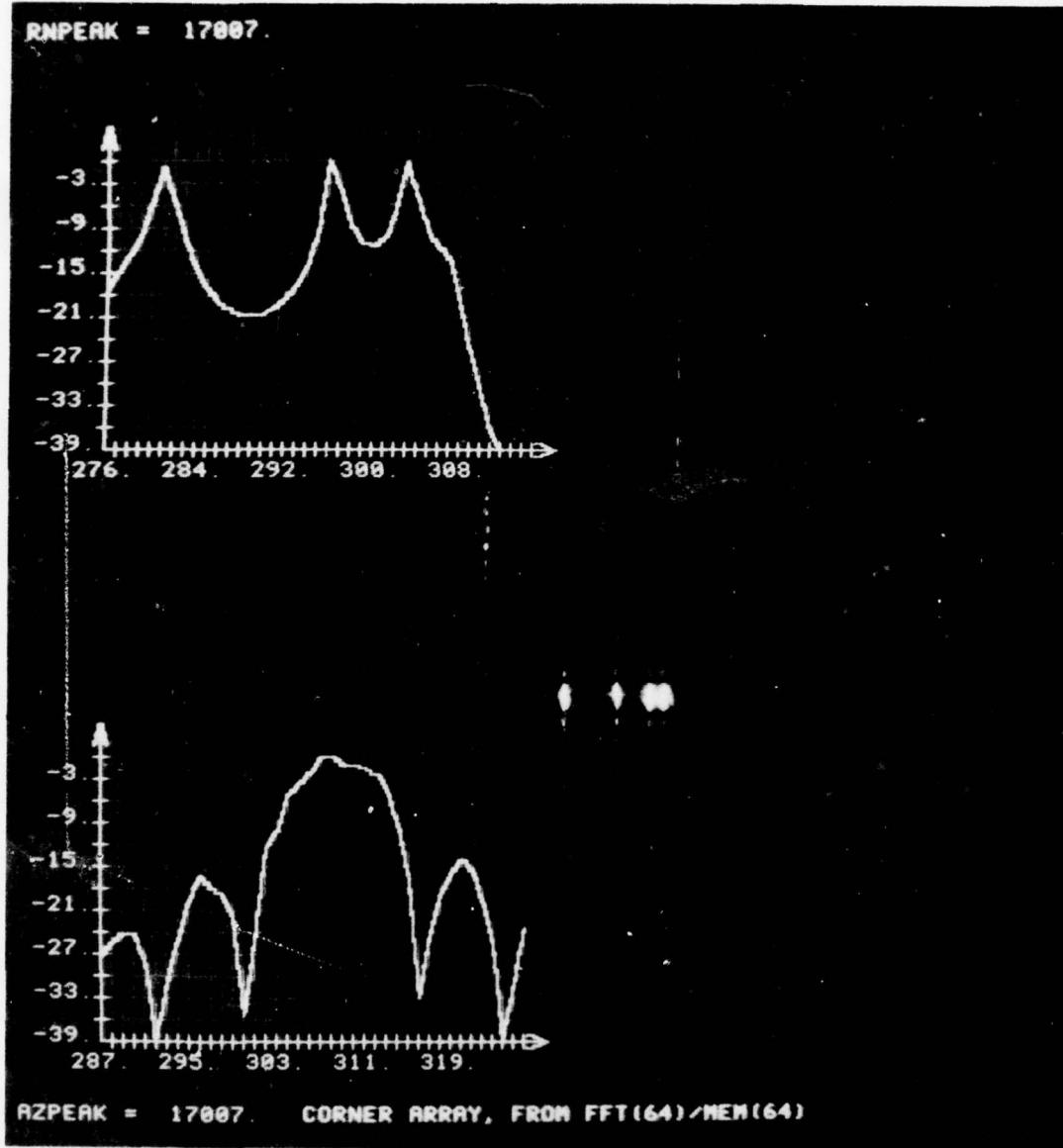


FIGURE 5. Intensity Modulated CRT Display of Horizontal MEM Processing and Vertical FFT Processing of  $64 \times 64$  Data Samples. Resolution is Retained Between Lower Right Reflector Pair. Upper Graph Shows Intensity Trace Across Three Lower Right Reflectors. Lower Graph Shows Vertical Intensity Trace Across Lower Right Reflector.

ANTENNA PATTERNS COMPUTED WITH MAXIMUM  
ENTROPY AND THE BURG TECHNIQUE <sup>1</sup>

WILLIAM R. KING

Consultant  
Alexandria, Virginia

Abstract

Antenna patterns computed with the technique of Maximum Entropy offer improved detection, beamwidth, and resolution. Example patterns are computed for omnidirectional spatial noise and complex receiver noise using 8 and 16 element one-dimensional antenna arrays.

Introduction

The Maximum Entropy Spectral Analysis (MESA) Technique as developed by John P. Burg [1] is applied to simulated spatial data sampled by M uniformly spaced elements of a one-dimensional antenna array. A spatial data series is processed with the MESA technique in the same way time series data is processed. However it is convenient to adopt complex number notation in representing signals incident to individual antenna elements, and consequently it is helpful to utilize a complex MESA algorithm having complex prediction error coefficients. The complex formulation of the Maximum Entropy has been described by Smylie, Clark and Ulrych [2] and Barnard [3].

Signals incident to the antenna are assumed to be spatially coherent plane waves with real and imaginary sinusoidal components. Spatial noise is assumed to be omnidirectional (or white noise) and is represented by a Gaussian amplitude distribution at each antenna element. However in some examples receiver noise is simulated by a random complex number in each of the N channels. Receiver noise is the most troublesome, since it is a complex vector and its presence distorts both signal phase and amplitude.

---

<sup>1</sup>This work supported in part by Rome Air Development Center, Contract No. F30602-75-C-0121 and Naval Research Laboratory, excepted appointment.

### Discussion

The expression for wavenumber power spectra is given by equation (1) in Figure 1, where it is noted that the prediction filter  $\gamma_n^N$  is a whitening filter such that the prediction error power  $P_n$  is a constant.<sup>n</sup> There are N filter coefficients where N is less than the number of data samples. The wavenumber component k is taken in the direction of the linear antenna so that it is a function of signal angle  $\theta$ . The antenna elements are spaced at half wavelength intervals,  $\Delta x$ .

The error power  $P_n$  and the filter coefficients  $\gamma_n^N$  are given by the recursive relations denoted in Equations (2) and (3). The last (N + 1) filter coefficient is a function of the backward and forward prediction errors given by  $\alpha_n^N$  and  $\beta_n^N$  respectively. The prediction errors  $\alpha_n^N$  and  $\beta_n^N$  are computed directly from the data samples. The set of equations in Figure 1 comprise the Burg technique which was introduced by Burg [4] in 1968.

A single signal incident at -3 degrees to the normal of an 8 element antenna is shown detected in Figure 2 by MESA for a (S/N) value of -5dB. For this and other examples (except where denoted) omnidirectional spatial noise is considered to be the dominant noise component. The signal is detected in a MESA snapshot and in an average of 20 such snapshots. A snapshot represents data which is recorded at one instant of time and processed with MESA. Snapshot patterns vary significantly when processing data recorded with short antennas. Consequently many patterns should be averaged to obtain a stable representative antenna pattern. Usually an average of 10 to 20 patterns provides good stability. Single signals are readily detected in omnidirectional noise even at low (S/N) values as indicated by Figure 2. Also the beamwidth and accuracy of a single MESA peak are quite good in omnidirectional noise.

In each figure the number of antenna elements NE, filter size N, and the random number generator seed no. IR are indicated as shown in Figure 2.

The problem of beam splitting is illustrated in Figure 3 where three signals are present at angles of -7, 0, and 7 degrees. The central signal at 0 degrees is split in the single MESA snapshot. However the same three signals are observed more accurately as three peaks in Figure 4 where 20 such snapshots are averaged. There is an apparent interaction between the three closely spaced signals as indicated by the shallow nulls between signals, and such interactions serve to limit the resolution possible with MESA.

Individual MESA snapshots of three signals located at -7, 0, and 7 degrees are shown in Figure 5 for (S/N) values of 13 dB and 40 dB in omnidirectional noise. Resolution could not be achieved at lower (S/N) values or for signals spaced closer together. Of course the three signals spaced 7 degrees apart are not resolved with conventional phased array beam summation as indicated

by the conventional pattern also shown in Figure 5.

The same three signals are shown just resolved in the MESA snapshot of Figure 6 for a (S/N) of 20 dB in complex receiver noise. In the same figure better resolution is indicated for a (S/N) value of 40 dB.

A MESA snapshot of five mixed signals of the same (zero) initial phase is shown in Figure 7 for dominant omnidirectional noise. The (S/N) values of each signal are indicated to vary from 0 to 20 dB. The stronger signals are readily detected whereas the weaker signal is observed only a few dB's above the MESA side peaks. All signal peaks are estimated by MESA within one degree of the true angle of incidence. However it is noted that the signal at 0 degrees is split into a strong and weak component. One remarkable characteristic of MESA is that the narrow beamwidth is maintained even at angles near 90 degrees.

All five signals are defined with even more precision in an average of 20 MESA snapshots as shown in Figure 8. Both precision and detectability are improved significantly by averaging. The split peak near 0 degrees is reduced to a very narrow single peak with pattern averaging. The conventional antenna pattern is shown for comparison, and it is observed that a signal beamwidth is considerably narrower in patterns computed with MESA. The weaker signal at 30 degrees is located more precisely with MESA, and of course the strong signal at 80 degrees is very poorly defined by the conventional pattern.

Phase shifts are investigated in the following three figures. Fougere [5] and Ulrych and Clayton [6] have noted that signal distortion occurs with use of MESA and the Burg technique when signals have a non-zero initial phase. And such distortion is most severe for a phase of 90 degrees. Phase shift distortion is demonstrated using the same five signals of the previous examples, where the strong signal at 60 degrees is given an initial phase of 90 degrees. The computed MESA antenna pattern snapshot is shown in Figure 9a where it is observed that the signals at -45, 30, and 60 degrees have all split into two components. However such splitting is virtually eliminated, and the stronger signals are more precisely defined in an average of 20 such MESA snapshots shown in Figure 9b. However the weaker signal at 30 degrees does not show much improvement in the averaged MESA pattern.

The same five signals are again detected with an averaged MESA antenna pattern shown in Figure 10, but the signal at 60 degrees has an initial phase of 180 degrees. Peaks are more severely distorted and broadened than in the previous example, and some peak splitting remains even after pattern averaging.

The effect of signal phase upon the resolution of two closely spaced signals is observed in the following two figures. Two 13 dB signals located at -3 and +3 degrees are shown resolved in the MESA snapshot of Figure 11.

The signal at +3 degrees is given an initial phase of 90 degrees and the resulting averaged MESA pattern is shown in the same figure. As a result of the initial phase shift, the two signals are shifted further apart by several degrees from their actual separation.

The same two signals are again observed in an average of 20 MESA snapshots shown in Figure 12 where the signal at +3 degrees has an initial phase of 180 degrees. Distortion and peak splitting is observed to be the most severe in the averaged pattern such that the two signals are both split into several peaks.

The conclusions resulting from this demonstration of MESA antenna patterns are summarized in Figure 13.

#### References

1. Burg, John P., 31 Oct. 1967, "Maximum Entropy Spectral Analysis," 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma.
2. Smylie, Clark and Ulrych, 1973, "Analysis of Irregularities in the Earth's Rotation," Methods in Computational Physics, Vol. 13, Academic Press, New York.
3. Barnard, T.E., 25 June, 1975, "The Maximum Entropy Spectrum and the Burg Technique," Tech. rept. no. 1, Advanced Signal Processing, ALEX(03)-TR-75-01, Texas Instruments Inc., Dallas, Texas.
4. Burg, John P. 1968, "A New Analysis Technique for Time Series Data," NATO Advanced Study Institute on Signal Processing With Emphasis on Underwater Acoustics, Enschede, Netherlands.
5. Fougere, Paul F., Dec. 1975, "A Solution to the Problem of Spontaneous Time Splitting in Maximum Entropy Power Spectrum Analysis," Annual AGU Meeting, San Francisco, California (EOS 56, 1054).
6. Ulrych and Clayton, 1976, "Time Series Modeling and Maximum Entropy," Phys-Earth Planet. Inter., Vol. 12, pp. 188-199.

$$P(k) = \frac{P_{N/K}}{\left| 1 + \sum_{n=1}^N \gamma_n e^{ikn(\Delta x)} \right|^2} \quad (1)$$

$$k = 2\pi/\lambda, \quad k = K \sin\theta, \quad \Delta x = \lambda/2$$

WAVENUMBER POWER SPECTRA

$$P_1 = r_o^2 \quad (2)$$

$$P_{N+1} = P_N \left[ 1 + (\gamma_{N+1}^{N+1})^2 \right]$$

$$\frac{\text{ERROR POWER}}{\text{POWER}} P_N$$

$$\gamma_1^N = 1$$

$$\gamma_{N+1}^{N+1} = \frac{2 \sum_{j=1}^{M-N+1} \beta_j^N \alpha_{j+n}^N}{\sum_{j=1}^{M-N-1} \left[ (\beta_j^N)^2 + (\alpha_{j+N}^N)^2 \right]} \quad (3)$$

$$\gamma_n^{N+1} = \gamma_n^N + \gamma_{N+1}^{N+1} \cdot \gamma_{N-n+2}^N$$

PREDICTION ERROR COEFFICIENTS

FIGURE 1. MAXIMUM ENTROPY AND THE BURG TECHNIQUE

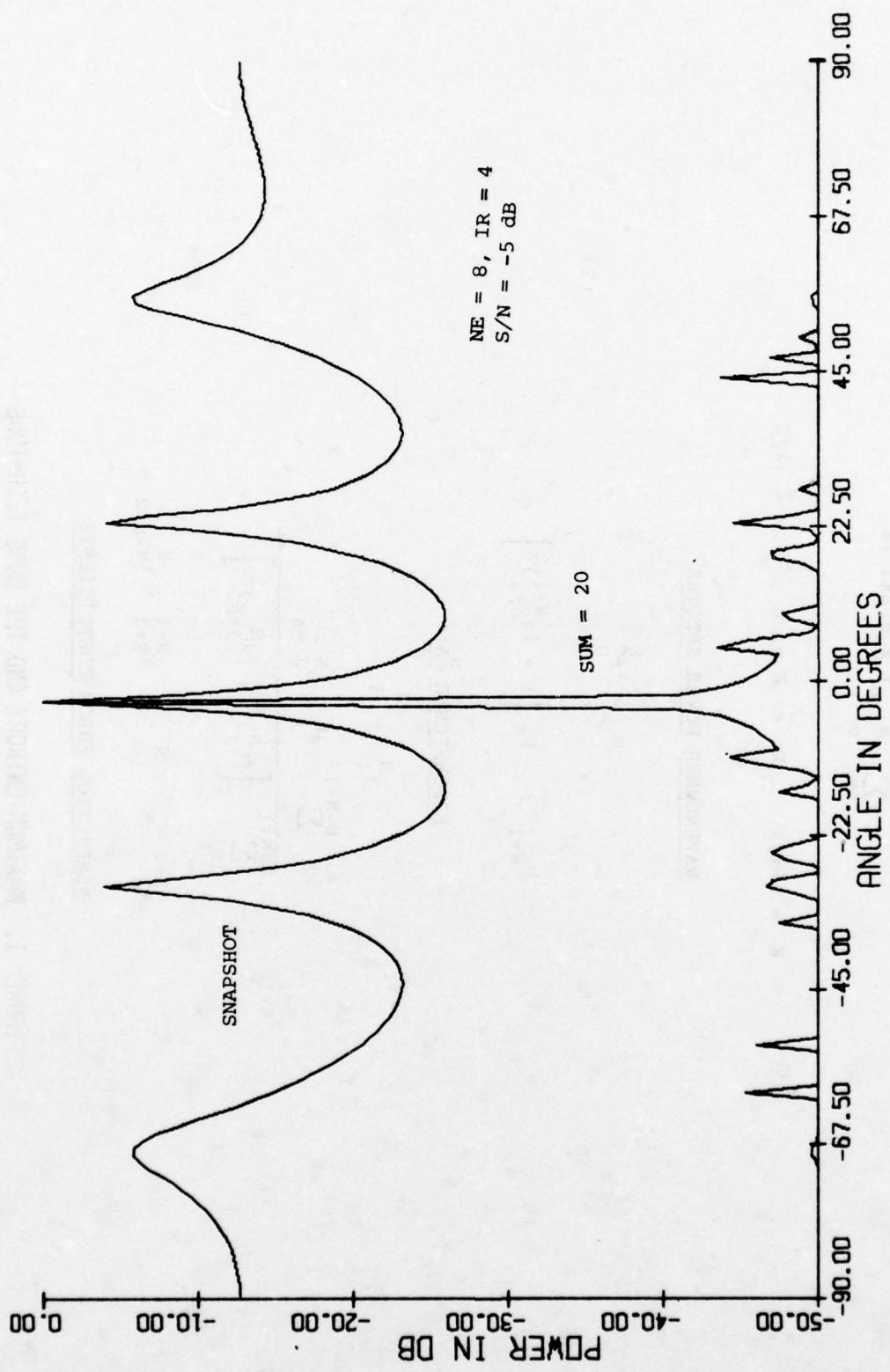


FIGURE 2. Signal at -3 Degrees

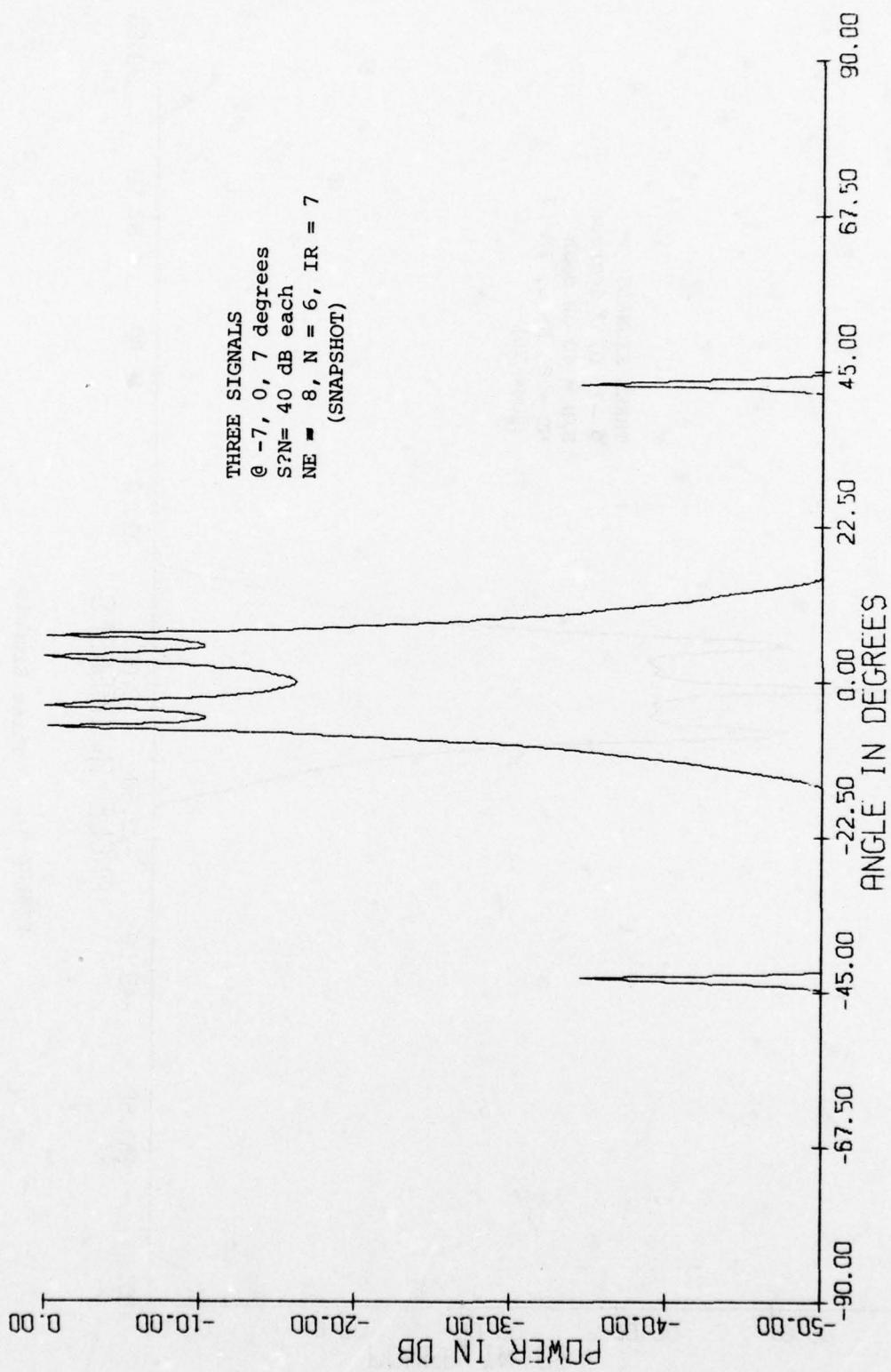


FIGURE 3. Beamsplitting with 3 Signals

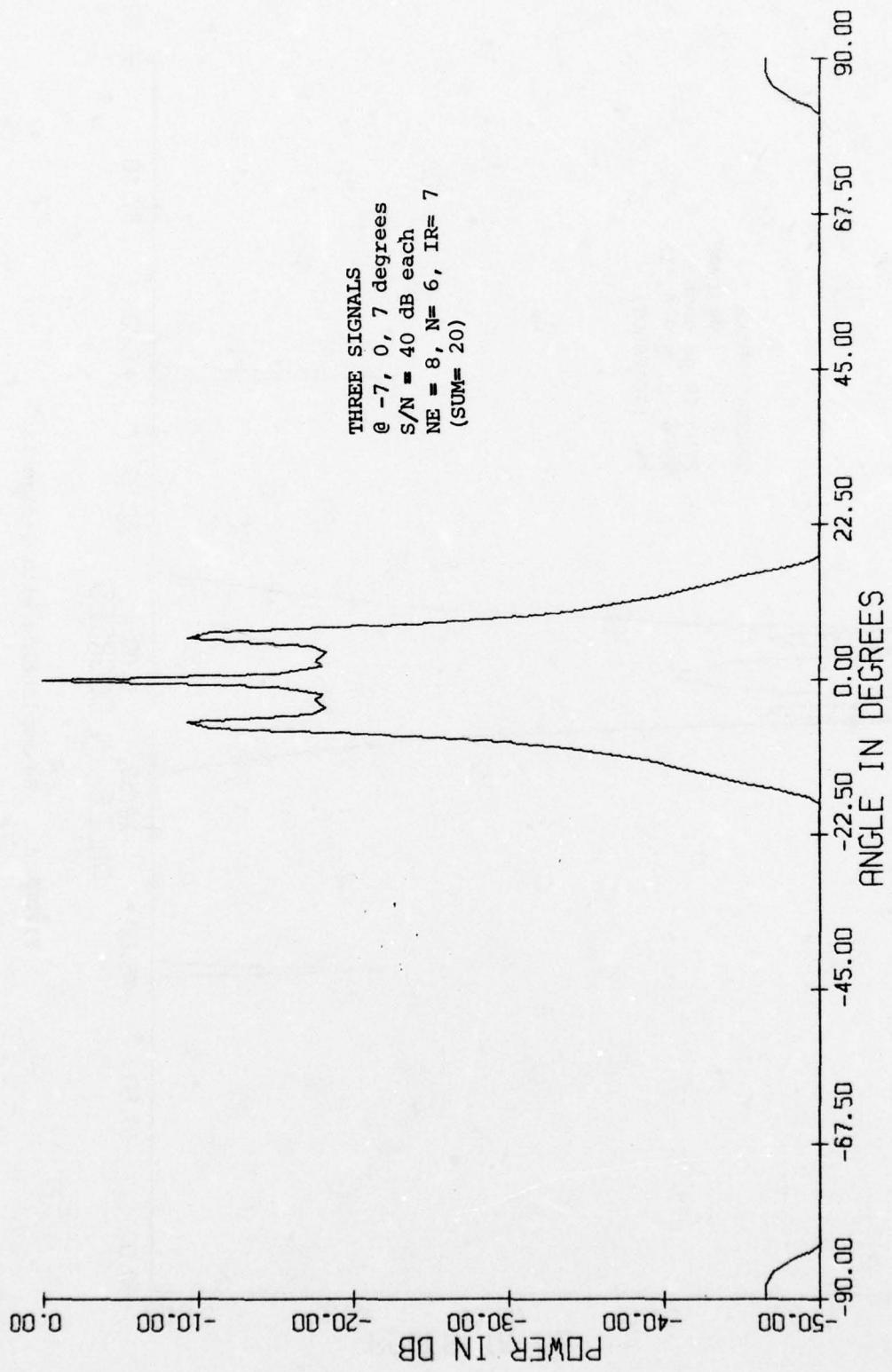


FIGURE 4. Three Signals

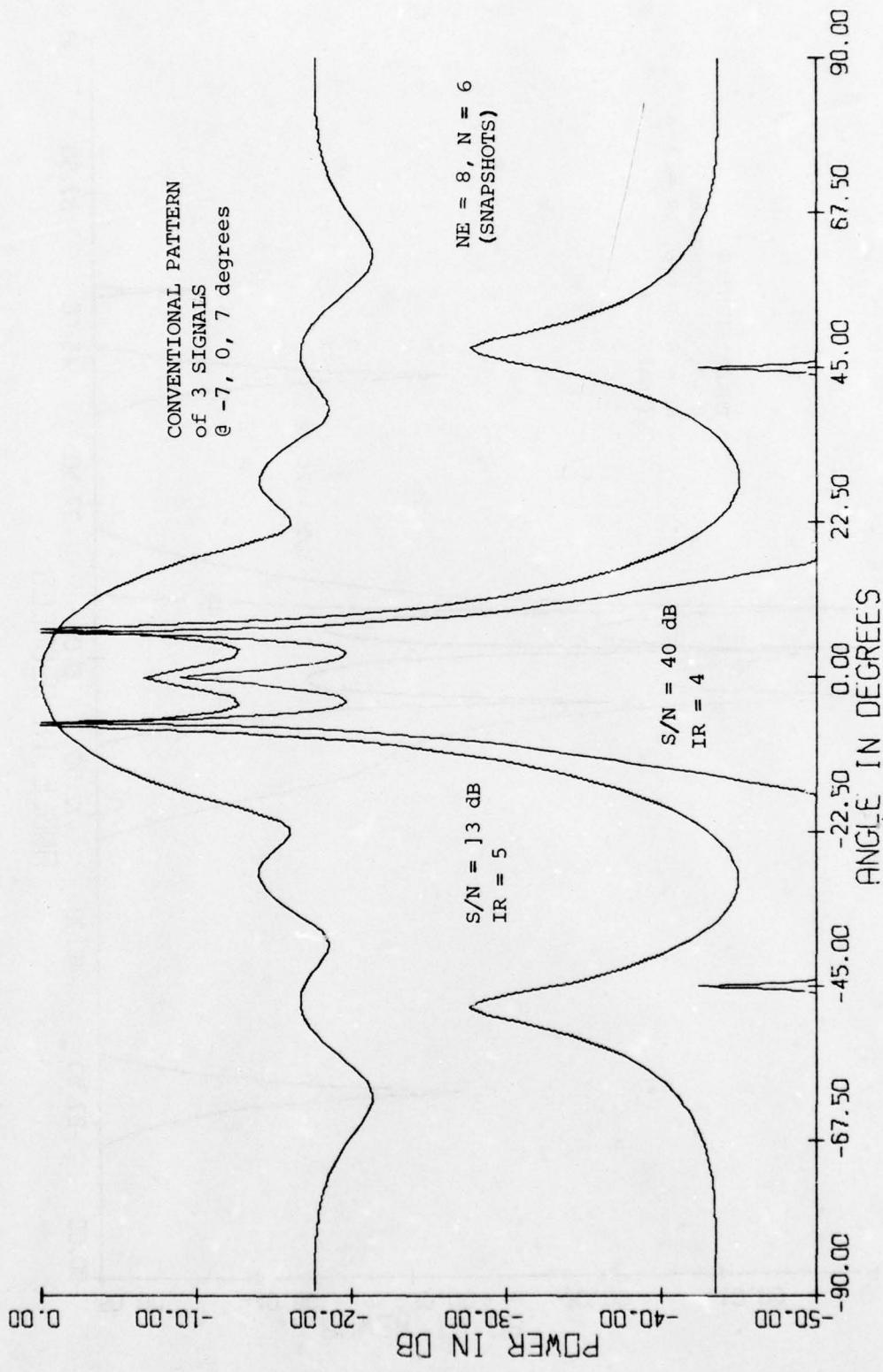


FIGURE 5. Resolution with Omnidirectional Noise

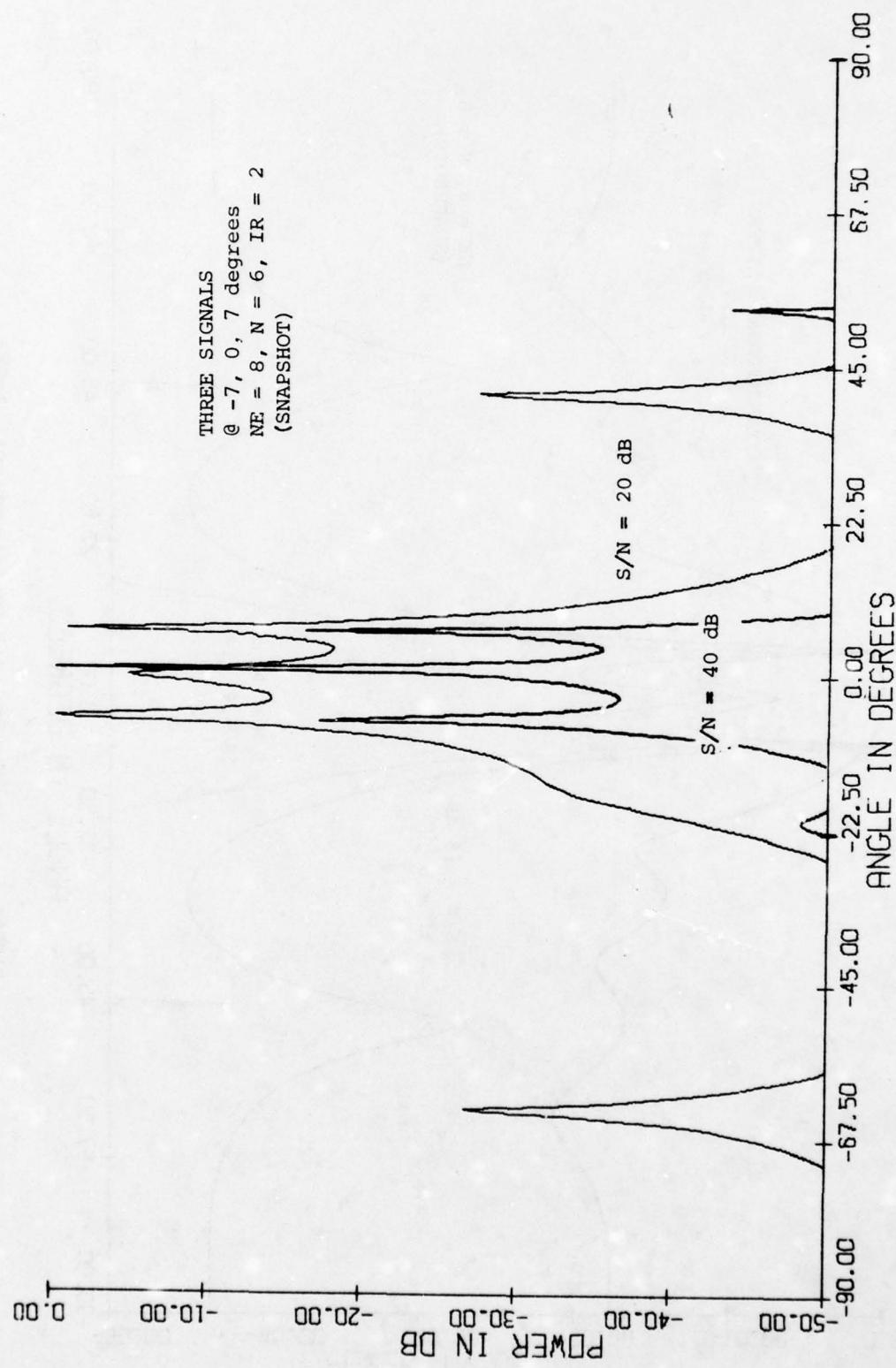


FIGURE 6. Resolution with Complex Noise

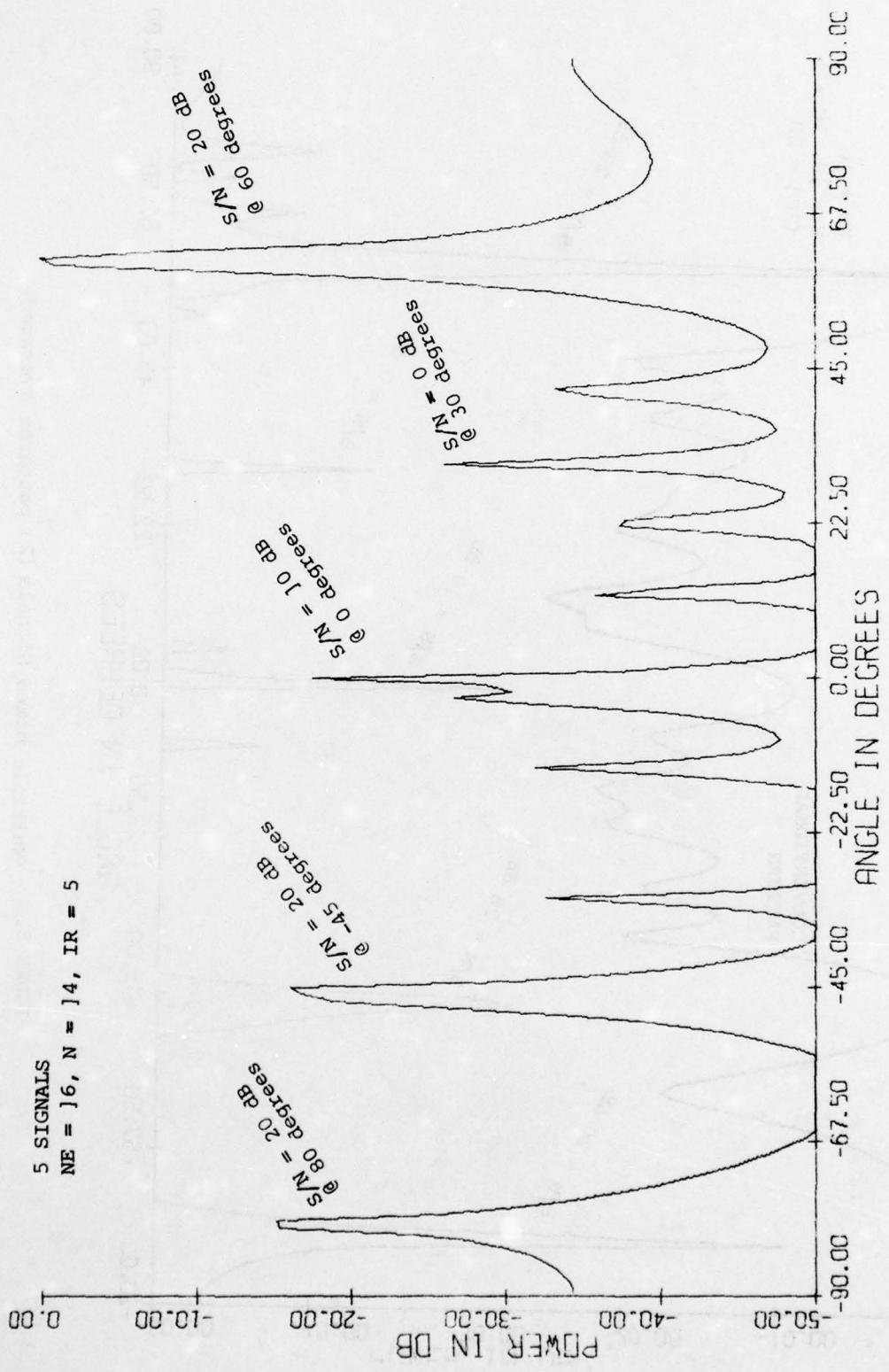


FIGURE 7. Multiple, Mixed Signals (snapshot)

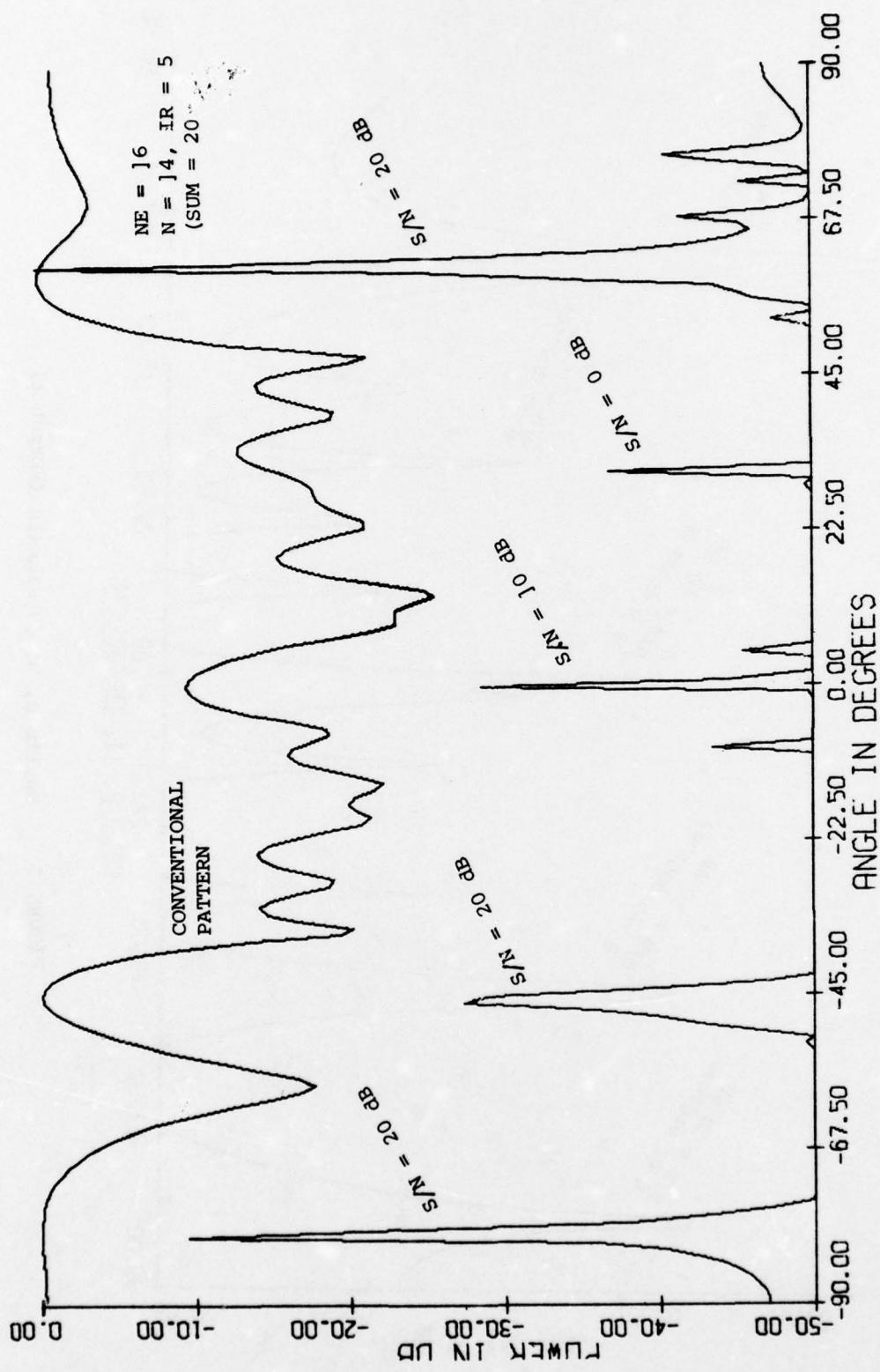


FIGURE 8. Multiple Mixed Signals (20 patterns averaged)

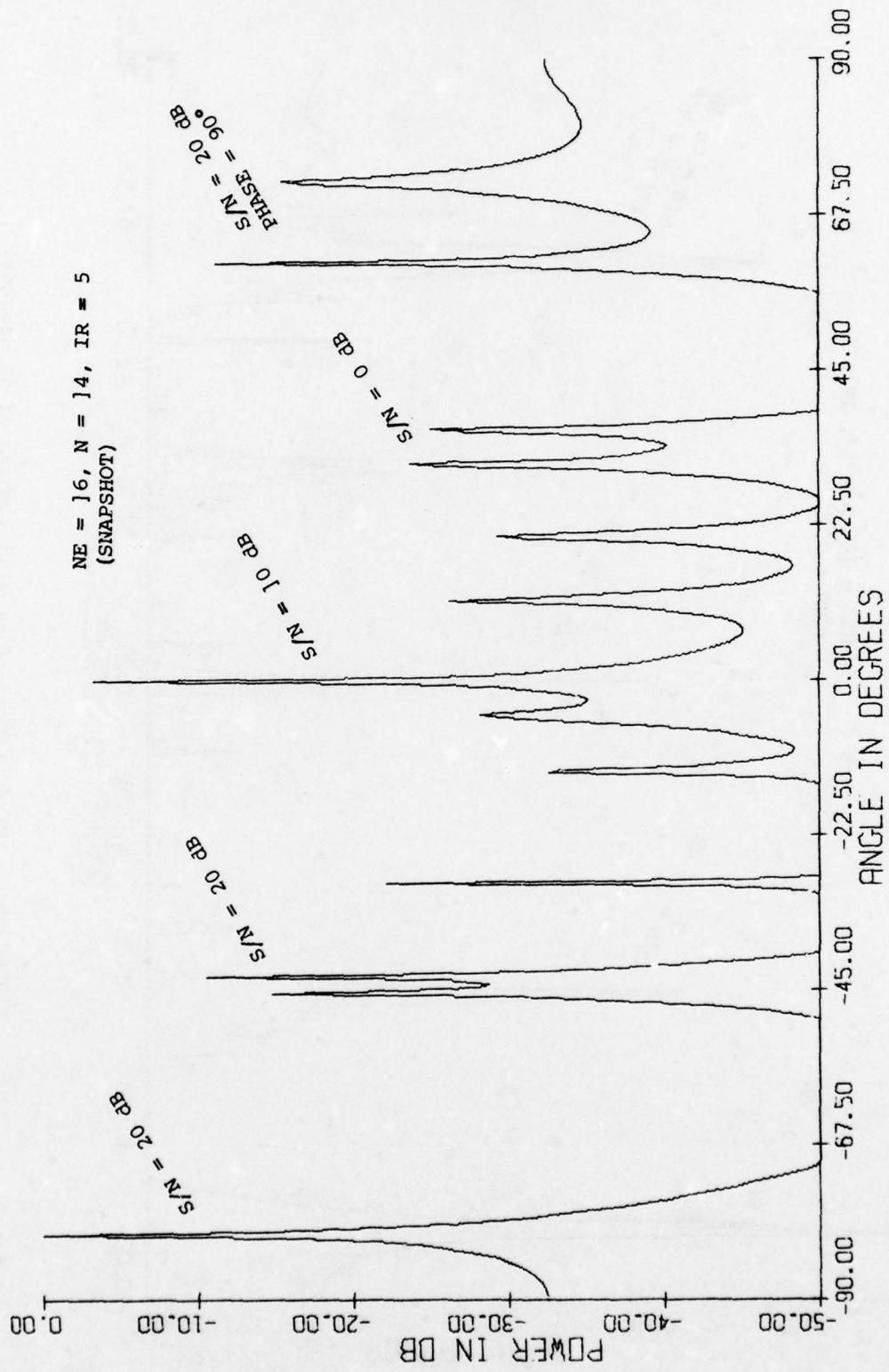


FIGURE 9a Mixed Signals (one with a phase of 90 degrees)

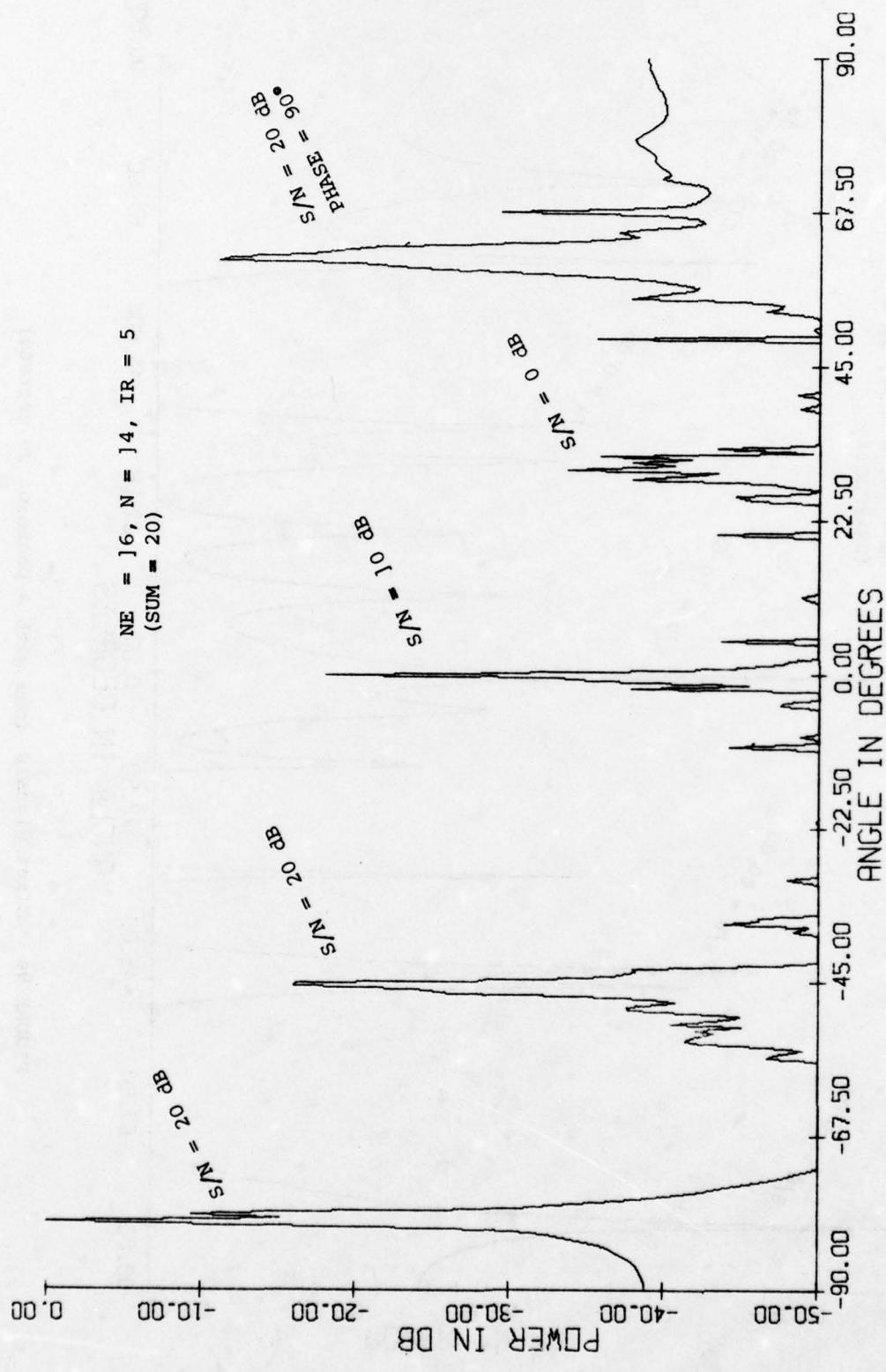
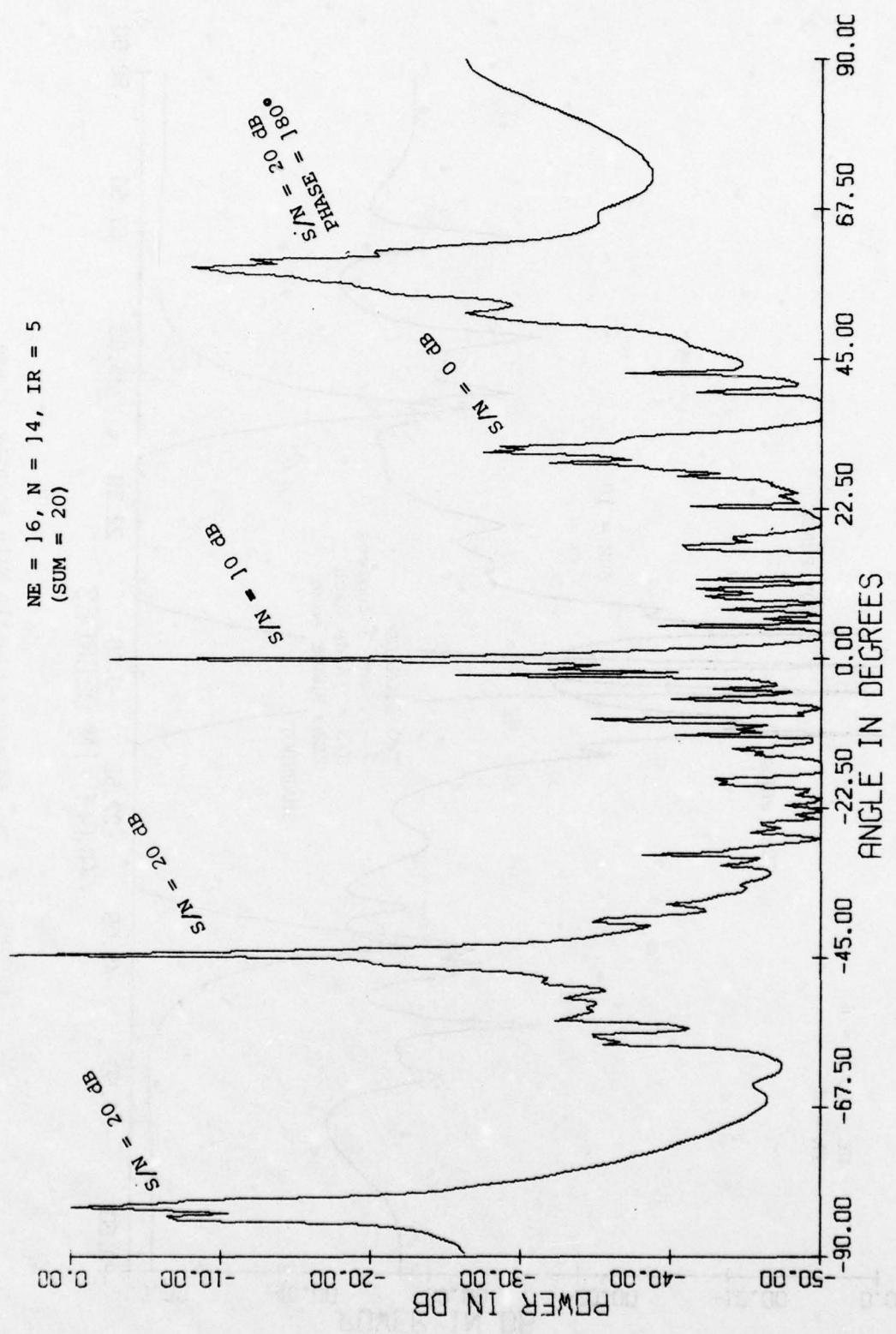


FIGURE 9b Mixed Signals (one with a phase of 90 degrees)



**FIGURE 10.** Mixed Phases (one with a phase of 180 degrees)

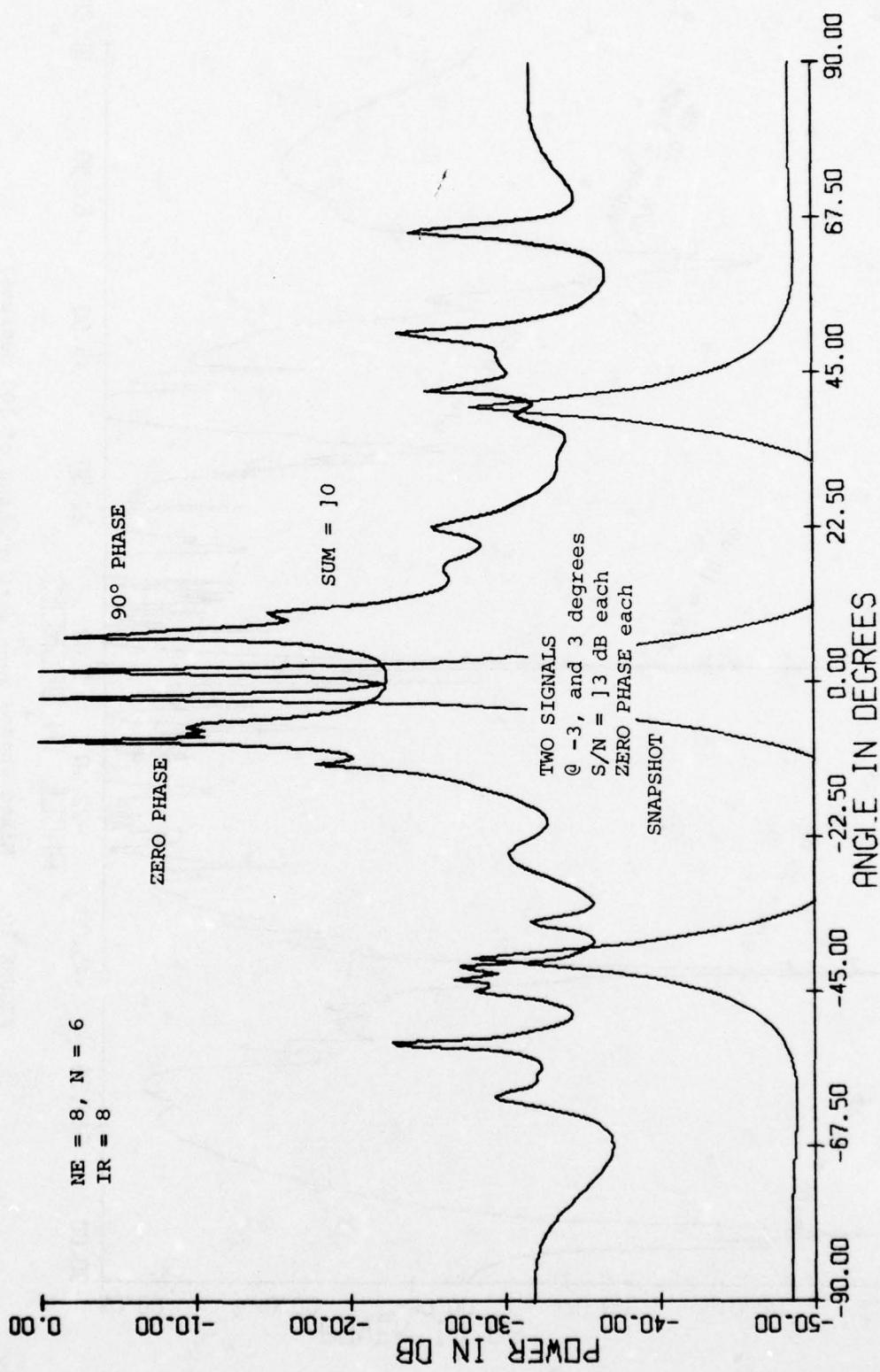


FIGURE 11. Two Adjacent Signals with Relative Phase

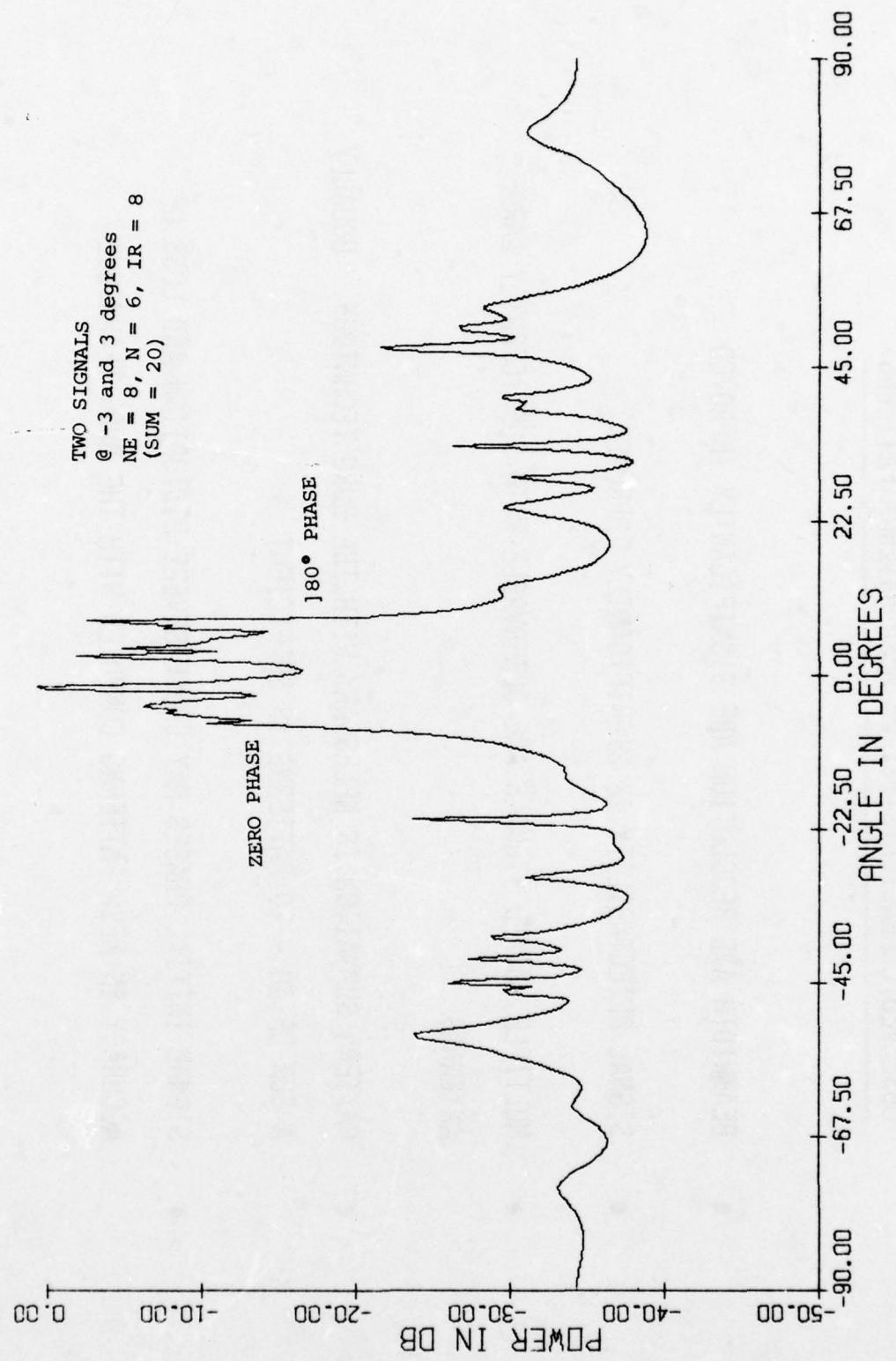


FIGURE 12. Two Adjacent Signals with Relative Phase

OBSERVED CHARACTERISTICS OF MESA ANTENNA PATTERNS

- BEAMWIDTH AND RESOLUTION ARE SIGNIFICANTLY IMPROVED
- SIGNAL DETECTABILITY IS SIGNIFICANTLY IMPROVED
- MULTIPLE, MIXED SIGNALS ARE OBSERVABLE WITH SUFFICIENTLY LARGE ANTENNAS
- PATTERN SUMMATION IS NECESSARY WITH THE BURG TECHNIQUE. USUALLY A SUM OF 10 - 20 PATTERNS IS SUFFICIENT
- SIGNAL INITIAL PHASES MAY CAUSE SEVERE DISTORTION AND LOSS OF ACCURACY IN MESA PATTERNS COMPUTED WITH THE BURG TECHNIQUE

FIGURE 13. CONCLUSIONS

## MAXIMUM ENTROPY CEPSTRAL ANALYSIS

TOM LANDERS

Applied Seismology Group  
Lincoln Laboratory  
Massachusetts Institute of Technology  
Cambridge, MA 02142

### Abstract

Cepstral analysis and maximum entropy spectral analysis have been combined and applied to the problem of identifying echos that arrive in the waveform of a primary signal. The method is referred to as maximum entropy cepstral analysis. By using Burg's technique for designing the complex prediction error spectral estimator, only that part of the complex log spectrum that lies in the band where the combination of source power and instrument bandpass produce good signal to noise ratios need be used to determine the cepstrum. By using only that data with good signal to noise ratios it is expected that more precise spectra than those obtainable by classical windowed spectral estimate methods will be obtained. The process order and consequently the filter length used to compute the maximum entropy cepstrum has been determined using Akaike's and Parzen's criteria as aids. On theoretical data with up to 50% background noise, echos at only a few digitizing intervals are detected. Used on a short period teleseismic recording of a seismic event where the echo time is known a priori, the technique finds the observed surface echo delay times.

### Introduction

Echo detection by cepstral analysis of band limited signals in the presence of broadband noise works poorly when the echo delay time is shorter than the primary signal. This situation results from the limitations of classical spectral techniques to resolve the delay harmonics in the complex log spectrum since the harmonics are embedded in noise whose amplitude varies logarithmically in the frequency band of interest. This paper deals with the application of a spectral estimation technique particularly suited to the analysis of such data, maximum entropy spectral analysis. We note that autoregressive spectral analysis and maximum entropy spectral analysis are one in the same.

When the current value of a time series is predictable with a white noise error from a finite linear combination of its past values, the process is called all-pole or autoregressive[1]. The spectrum of the time series is directly obtainable from the linear prediction coefficients and the power level of the prediction error series. From a practical point of view since any real set of data is bounded in time the technique used to estimate the set of prediction error coefficients must be selected and applied with care if meaningful spectral estimates are to be obtained.

The first section of this paper deals with the algorithm used to estimate the prediction error coefficients and methods used to determine the number of coefficients necessary to describe the model, called the order number of the process. Following sections deal with the application of the technique to echo detection. The first examples examine the effect of various signal to noise conditions and very short delay times using synthetic data. The paper concludes with the analysis of an actual piece of seismic data in which the echo delay time is known from measurements at the source.

#### Parameter Estimation

The time series for which we wish to have a spectral estimate is

$$x_t, t = -\infty, \dots, -1, 0, 1, \dots, \infty \quad (1)$$

for which we have the observed portion at  $t=1, \dots, N$ . The autoregressive model of  $x$  is

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + \epsilon_t \quad (2)$$

where  $\epsilon_t$ , called the innovation or prediction error, is, in our case, white noise. The spectrum is given by

$$S(f) = 2\Delta t E_p / \left| 1 - \sum_{j=1}^p A_j \exp(i2\pi f j \Delta t) \right|^2 \quad (3)$$

where  $E_p$  is the total residual square prediction error for prediction filter length  $p$ . Thus to calculate the power spectrum of  $x$  we must determine the order number  $p$  and the  $p$  values of the prediction filter.

Several approaches are available for estimating the prediction coefficients. The one we've chosen is due to Burg[3]. In this method the measure of the error in the construction of the coefficients is taken as the sum of the total squared prediction error resulting from predicting forward in time with the same quantity for predicting backward in time, a physically reasonable measure under the assumption that the data is stationary. Burg showed that if the prediction coefficients for order  $p$  were produced from a

linear combination of the prediction coefficients of order  $p-1$  and the reverse time prediction coefficients of order  $p-1$  such that the above measure of error is minimized by the choice of the scale factor of the combination, then the autocorrelation function will be necessarily positive definite. Use of the Burg algorithm has the side effect that the location of spectral peaks are somewhat dependent on the initial phase[2,4]. In practice, the degree of this error can be estimated by comparing spectral estimates of slightly shifted versions of the data set under consideration. While other methods for determining the prediction coefficients that have comparable resolution to the Burg method do not seem to have this problem they do not necessarily produce positive definite autocorrelation functions and are computationally less efficient. On this basis we used the Burg technique for parameter estimation.

It remains to determine the value of  $p$ , the order number. The following methods have been proposed. Akaike[5,6] suggested that the minimum of the average error due to estimating the autoregressive coefficients and the innovation for one step prediction gives  $p$ . The criterion, called the final prediction error, FPE, to be minimized is

$$FPE(M) = P_M(N+(M+1))/(N-(M+1)) \quad (4)$$

where  $P_M$  is the residual squared error for an  $M$ th length filter and  $N$  is the length of  $X$ . The  $M$  for which FPE is a minimum is taken as  $p$ . Alternatively Akaike[7,8] suggested minimizing the log likelihood of the innovation variance as a function of filter length to find  $p$ . This criterion, called the information theoretic criterion, AIC, is estimated by

$$AIC(M) = \ln(P_M) + 2M/N. \quad (5)$$

Again the  $M$  for which the AIC is minimized is taken as  $p$ . A third method considered here was proposed by Parzen[9,10] and is known as the autoregressive transfer function criterion, CAT. The order  $p$  is given where the estimate of the difference of the mean square error between the true filter, which exactly gives the innovation, and the estimated filter is a minimum. Parzen showed that this difference can be calculated, without explicitly knowing the exact infinite filter, by

$$CAT(M) = \frac{1}{N} \sum_{j=1}^M \hat{P}_j^{-1} - \hat{P}_M^{-1} \quad (6)$$

where

$$\hat{P}_j = N/(N-j)P_j.$$

In the examples that follow the Burg method is used to determine the  $A_i$ 's. For plotting purposes, noting that  $\log FPE$  asymptotically approaches AIC, we define

$$\begin{aligned}
 F(M) &= \log_{10} (\log_{10} (FPE(M)) - \log_{10} (FPE(p)) + 1) \\
 A(M) &= \log_{10} (AIC(M) - AIC(p) + 1) \\
 C(M) &= \log_{10} (CAT(M) - CAT(p) + 1)
 \end{aligned} \tag{7}$$

such that the value of each criterion at the predicted order number,  $p$ , is 0.

Landers and Lacoss[15] examined the method for complex sinusoids with various signal to noise ratios. Their results indicated that where little noise is present, AIC, FPE and CAT give an order number that produces an acceptable spectra. When added white noise was large, the criteria gave inconsistent values and, in these cases larger filter lengths produced better spectra. We note that it is often the location and size of narrow band components which are of scientific interest but that the various order criteria select filter orders based upon the entire spectrum including added white noise. For example although large filter lengths produced the best fit spectra, the complex harmonic  $\exp(\sqrt{-1}\omega_0 t)$  can be exactly predicted by the one point prediction error wavelet  $(1, -\exp(\sqrt{-1}\omega_0))$ .

For a single echo in a broadband primary, the complex log spectrum looks like

$$F(f) = \log(1 + \alpha e^{i2\pi ft}) \tag{8}$$

where  $t$  is the echo delay time and  $\alpha$  the echo size. Four cycles of this function for  $\alpha$  equal .5 and a delay time of 1 with 10% white noise added are plotted at the top of Figure 1. Note that frequency and time have been interchanged. While the function is all-pole, it has an infinite number of poles, that is to say an infinite order number. The error functions, shown below the time series, indicate that the filter length should be greater than twenty-three though no clear minimum value is discernable. The maximum entropy spectrum for a filter length of twenty-three is shown at the bottom, the dashed line indicating the locus line of the amplitudes of the exact spectral peaks. The computed amplitudes vary by a few dB from the line and up to the eighth order give the correct frequency. In general we use the FPE, AIC, and CAT criteria as guides to picking the order number reserving the possibility that larger or smaller values of  $p$  can provide useful information on the nature of the physical system being modeled. We now proceed to use the Burg algorithm, Akaike's and Parzen's order number criteria and maximum entropy spectral analysis to compute cepstra of time series containing echoes.

### Cepstral Analysis

Our examples deal with echo detection in seismic data. The nature of the first few seconds of a short period seismogram for a shallow event is determined by the shape of the wave that travels directly from the source region to the seismometer and the underside reflection off the earth's surface that results in the addition of a delayed and scaled version of the direct wave. The reflection is referred to as the depth phase since its time delay coupled with the known elastic velocity structure in the upper layers of the earth allows one to calculate the depth of the event. Since the amplitude of depth phases are often as large as the amplitude of the direct waves and, since short period seismic systems are usually quite narrow band, depth phases that arrive while energy from the primary arrival is still coming in cannot be clearly visually detected. Assuming that the secondary arrivals differ from the primary wave by only a scale factor and some dispersion due to anelastic attenuation, detection may in theory be accomplished by cepstral analysis[12]. Unfortunately, when our echo is only slightly delayed from the primary, the log spectrum over the band of good signal to noise ratio contains only a few cycles of the primary harmonic that peaks the cepstrum at the echo delay time. Under such conditions normal spectral analysis of the log spectrum will not produce well defined cepstra. Thus we employ maximum entropy spectral analysis to obtain the necessary resolution.

In Figure 2, a synthetic seismogram with echoes at 1 and 3 seconds in 5% white noise is shown at the top. The complex log spectra and the power spectra are shown below the seismogram. The nonlinear effects of noise can be seen in the regions of the spectra where the signal power is low. To minimize this effect, only the band from .5 to 2 hertz was used to compute the complex Burg filter coefficients. The error functions corresponding to these coefficients are shown at the bottom of the figure. The same functions were computed for signal to noise ratios in the time signal of .001, .05, .5 and 1, each time limiting the log spectral band to the region of good signal to noise. The maximum entropy cepstra for these cases are shown in Figure 3, for decreasing signal to noise ratios from top to bottom. While the lowest noise case gives the exact result, signal to noise ratios up to .5 give satisfactory results. Figure 4 shows the analysis for an echo delayed by only four digitizing intervals in 5% white noise. Though the delay time is not well resolved the peak occurs at the correct time.

The following illustrates the method as applied to a real seismogram recorded in Norway for a source in Nevada. It is known from instruments at the source site that the depth phase will be delayed from the initial arrival by .98 seconds. Referring to Figure 5, part (a) shows the recorded seismogram and part (b) the tapered version that isolates the first arrival and the depth phase. Taking the discrete fast Fourier transform of part (b) and then the analytic complex logarithm results in the power spectrum, real log spectrum and imaginary log spectrum shown in part (c). The effect

of the short period seismometer and anelastic attenuation has been compensated for in the log spectrum. Selecting the band over which there is sufficient power to insure that the behavior of the imaginary part of the log spectrum (the phase of the spectrum) is controlled by signal rather than noise conditions and removing linear trends from the real (a source correction) and imaginary (a simple time delay of the whole trace) produces the reliable part of the complex log spectrum shown in part (e). The power spectrum of this complex frequency function should have peaks at the delay and multiples of the delay time of the depth phase. Since the delay time is known to be near 1 sec the log spectrum should have a primary periodicity of about 1 hertz. The total band is only 1.5 hertz long so that we wish to make a spectral estimate of series that contains only about 1.5 cycles of the primary component. Applying the Burg algorithm to this complex frequency function and calculating the Akaike FPE and AIC functions from the error that results when the Burg coefficients are applied, we obtain the curves shown in part (f). The model parameters, namely the prediction coefficients, for the order corresponding to the minimum in the order number curves are shown in part (g). The exact spectrum of the model which is the estimated spectrum of the log spectrum or cepstrum is shown in part (h). Detected in the cepstrum is the echo delay time at .95 and its first harmonic at 1.9 ( $2 \times .95$ ) seconds. The predicted time delay between the first peak of the primary arrival and the first peak of the depth phase, which should be reversed in sign, is shown in part (i). The difference between the known and predicted delay time is less than 1 digitizing unit, the original data being sampled at 20 hertz. The same technique applied[13,14] to other sources and multiple recordings of single sources shows that for delays of more than approximately .5 seconds that the depth phase delay time falls within approximately .1 seconds of the correct time. In many cases the seismograms did not show any obvious change of character at the delay time.

#### Discussion

The success of maximum entropy or autoregressive spectral analysis in large part depends on how well one is able to model a process. In our examples, where the physical nature of the phenomena dictated that all-pole or prediction error models were appropriate, we were able to obtain useful spectra. By using the Burg algorithm, model coefficients were produced without making any unrealistic assumptions about the data outside the actual observation interval. In each case, spectra consistent with known data properties or internally consistent among independent data sets confirmed that the technique was producing reliable results. While order number criteria are not perfect and problems exist in determining model parameters the technique can provide useful results in cases where classical estimation techniques are not applicable.

### Acknowledgements

This work was sponsored by the Advanced Research Projects Agency of the Department of Defense. The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

### References

1. Makhoul, J., 1975 "Linear Prediction: A Tutorial Review", Proc. IEEE, 63, 561-580.
2. Ulrych, T. J. and Clayton, R. W., 1976, "Time Series Modelling and Maximum Entropy", Phys. Earth and Planet. Inter., 12, 188-200.
3. Burg, J. P., 1975, "Maximum Entropy Spectral Analysis", Ph.D. Thesis Stanford University, Palo Alto, CA, 123pp.
4. Chen, W. Y. and Stegen, G. R., 1974, "Experiments with Maximum Entropy Power Spectra of Sinusoids", J. Geophys. Res., 79, 3019-3022.
5. Akaike, H., 1969, "Power Spectrum Estimation Through Autoregressive Model Fitting", Ann. Inst. Stat. Math., 21, 407-419.
6. Akaike, H., 1970, "Statistical Predictor Identification", Ann. Inst. Stat. Math., 22, 203-217.
7. Akaike, H., 1972, "Use of an Information Theoretic Quantity for Statistical Model Identification", Proc. 5th Hawaii Int. Conf. on System Sciences, 99-101.
8. Akaike, H., 1974, "A New Look at the Statistical Model Identification", IEEE Trans. Automat. Control, AC-19, 716-723.
9. Parzen, E., 1976, "An Approach to Time Series Modelling and Forcasting Illustrated by Hourly Electricity Demands", Tech. Rpt. No. 27, Statistical Science Division, State University of New York at Buffalo.
10. Parzen, E., 1976, "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes", Tech. Rpt. No. 23, Statistical Sciences Division, State University of New York at Buffalo.
11. Landers, T. E., 1974, "Improved Depth-Phase Identification via Maximum Entropy Cepstral Analysis", Seismic Discrimination Semiannual Technical Summary, Lincoln Laboratory M.I.T., 26-30, 30 June.

12. Bogert, B. P., Healy, M. J. R., and Tukey, J. W., 1963, "The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking", Proc. of the Symposium on Time Series Analysis, ed. M. Rosenblatt, Wiley and Sons, Inc., New York, 206-243.
13. Landers, T. E., 1975, "Maximum Entropy Cepstral Analysis", Seismic Discrimination Semiannual Technical Summary, Lincoln Laboratory M.I.T., 1-5, 30 June.
14. Landers, T. E., 1975, "Studies on the Maximum Entropy Spectral Technique" Seismic Discrimination Semiannual Technical Summary, Lincoln Laboratory M.I.T., 49-67, 31 December.
15. Landers, T. C., and Lacoss, R. T., 1977, "Some Geophysical Applications of Autoregressive Spectral Estimates", IEEE Trans. on Geosci. Elect., GE-15, 26-31.

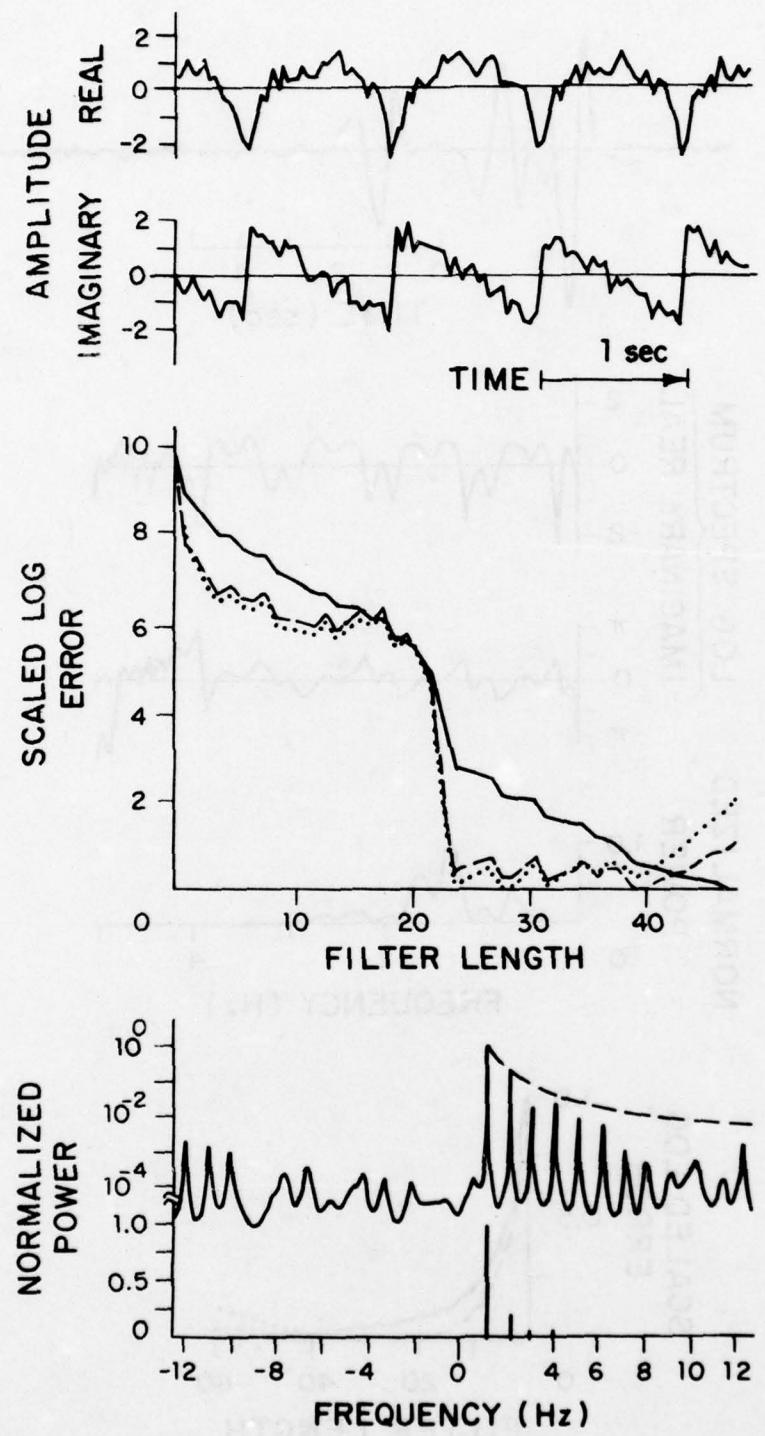


FIGURE 1. Error analysis for  $\log(1 + 0.5 \exp(\sqrt{-1} 2\pi t))$

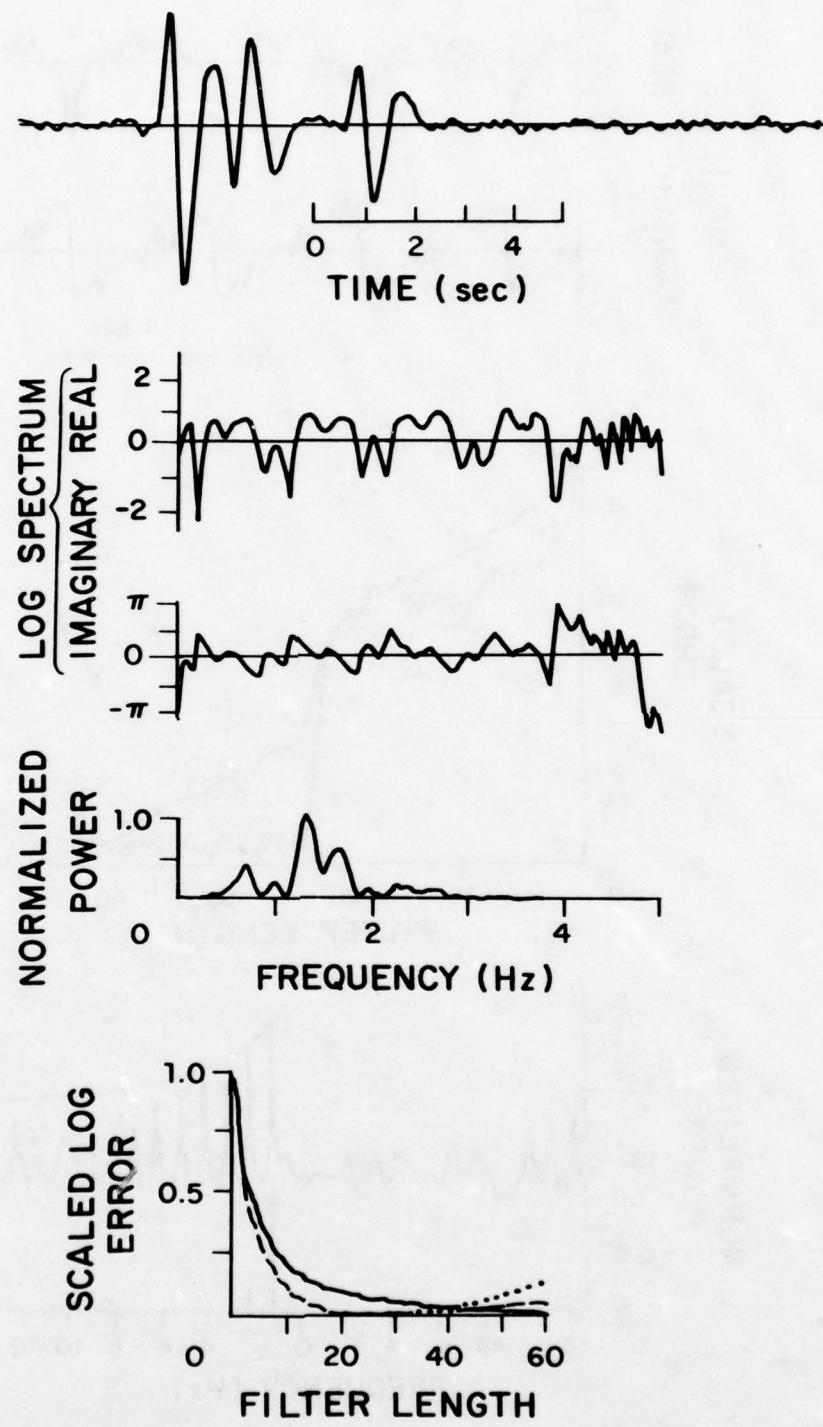


FIGURE 2. Error analysis of the log spectrum of a time function with echoes at 1 and 3 seconds.

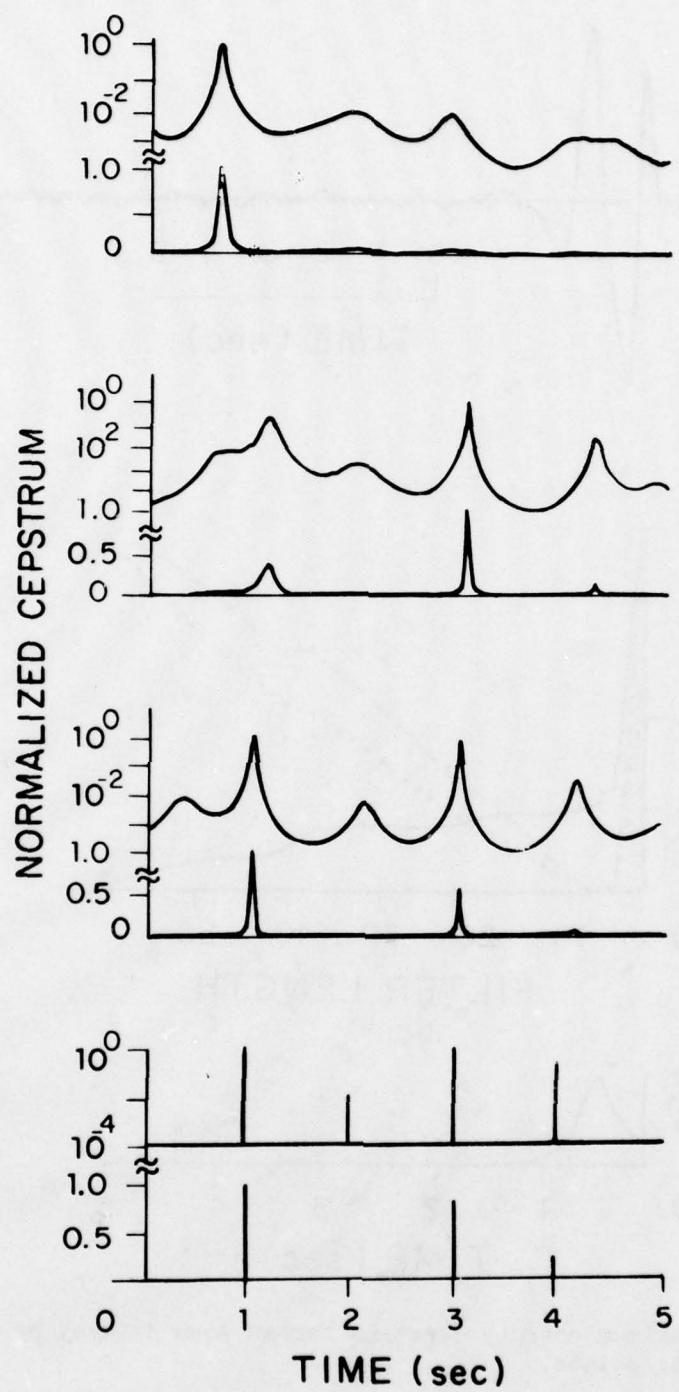


FIGURE 3. Maximum entropy cepstra for the time function given in Figure 2 with signal to noise ratios of (from top to bottom) 1, .5, .05, .001.

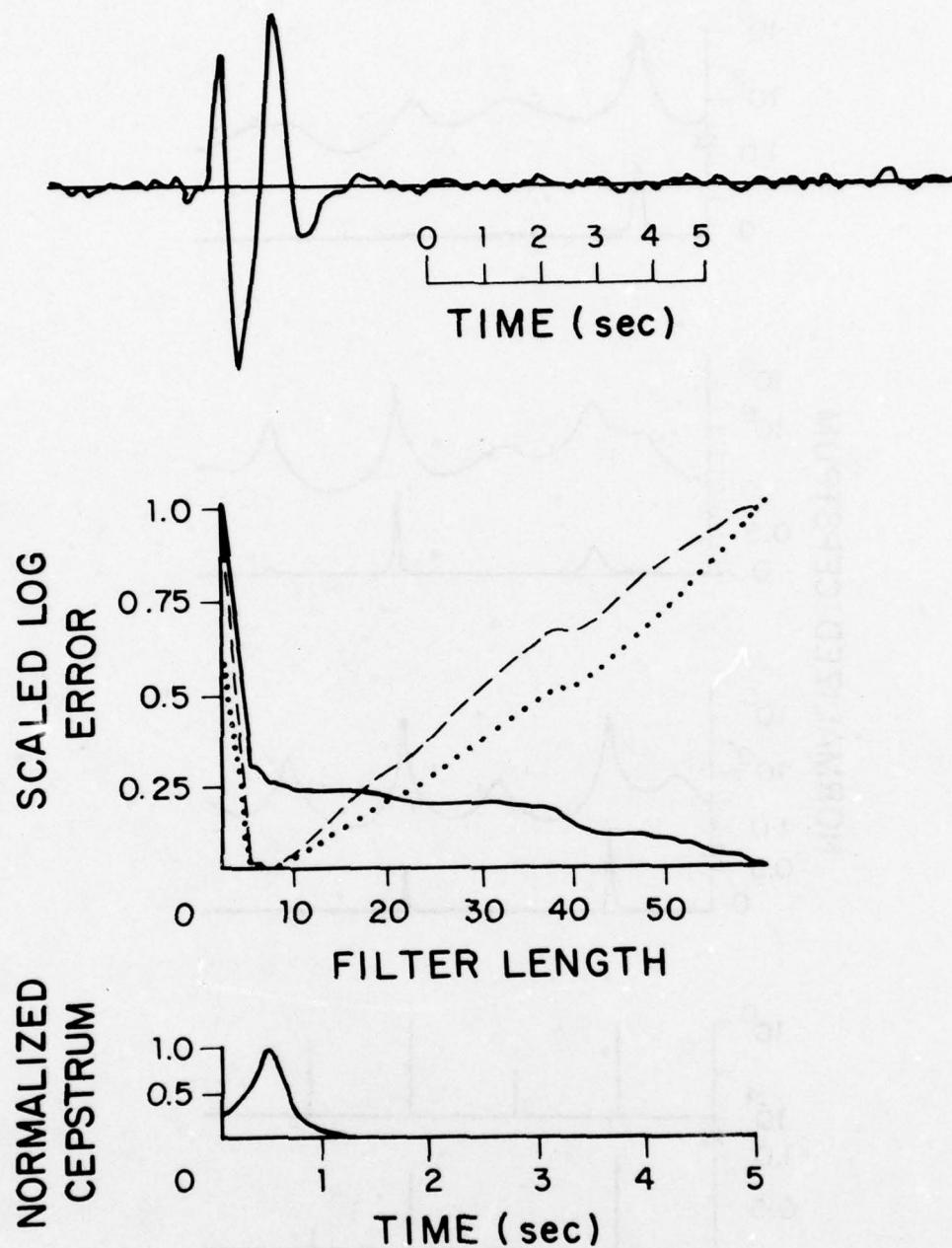


FIGURE 4. Maximum entropy cepstrum for an echo delayed by 4 digitizing points.

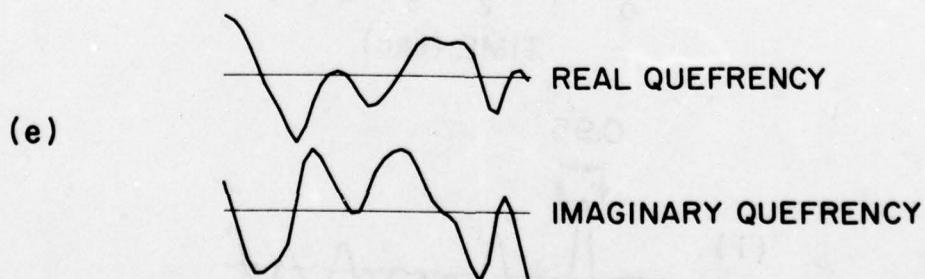
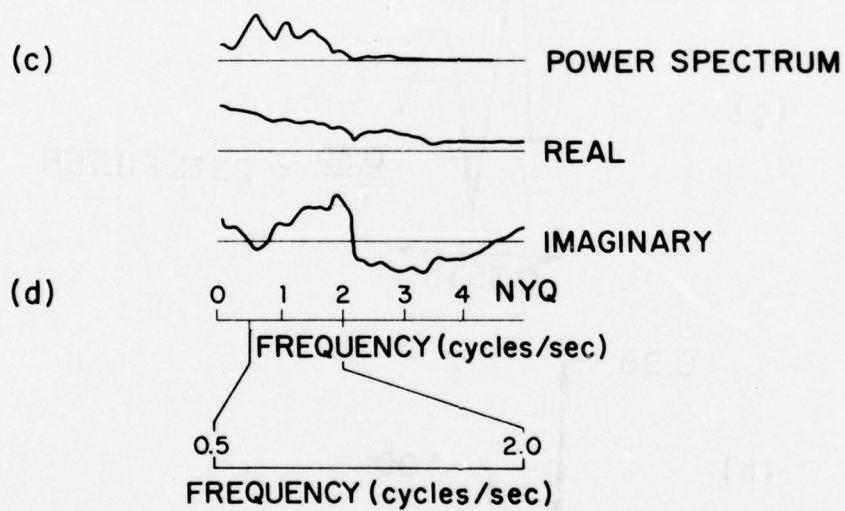
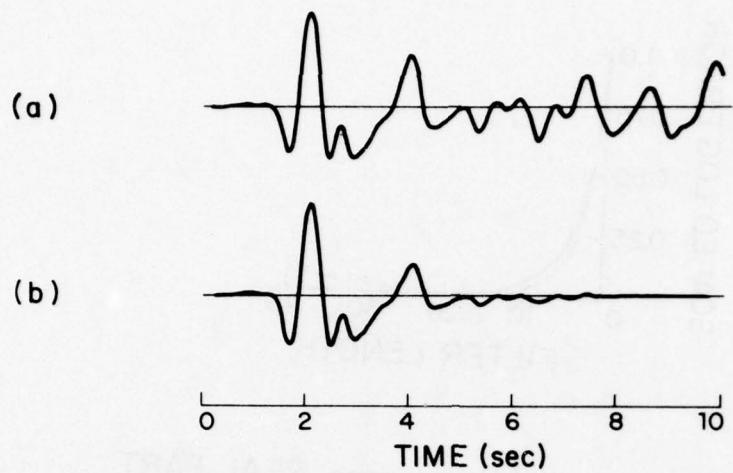


FIGURE 5. Maximum entropy cepstral analysis for an event with an echo known to arrive at about .98 seconds.

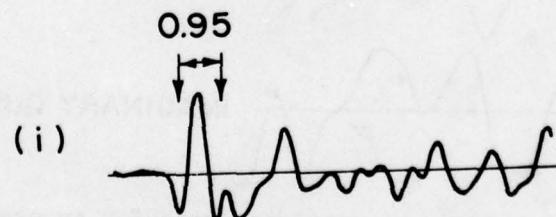
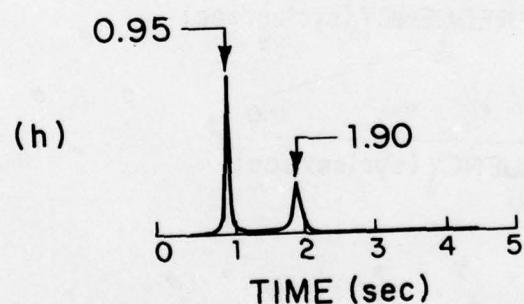
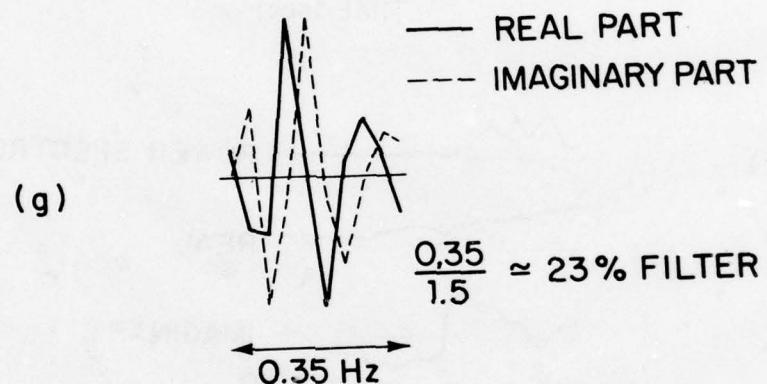
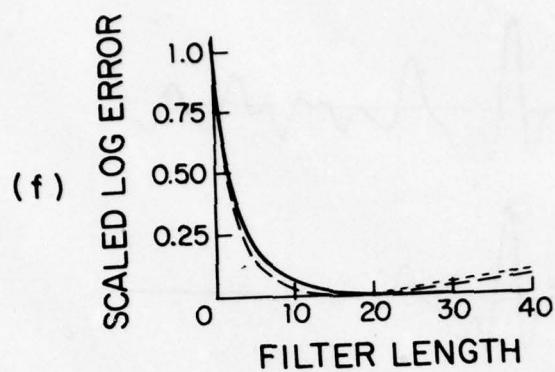


FIGURE 5. Continued.

# A NEW ADAPTIVE FILTER FOR RADAR CLUTTER REJECTION

D. E. BOWYER  
P. K. RAJASEKARAN  
W. W. GEBHART

Radar Systems Department  
Teledyne Brown Engineering  
Huntsville, Alabama 35807

## Abstract

A new adaptive filter to reject clutter is derived using autoregressive spectral analysis techniques. The adaptive filter performs open loop processing, resulting in a shorter transient response, and is therefore suitable for radar waveforms containing only a small number of samples. A number of examples including application to Ballistic Missile Defense are presented to demonstrate the performance capabilities of the new adaptive filter.

## 1. Introduction

This paper presents a new adaptive filter (AF) for clutter suppression, utilizing an autoregressive (AR) spectral analysis technique. Such a filter can be useful in tactical situations where the clutter environment is changing, such as in air defense (chaff), airborne moving target indication (MTI), and tank break-up clutter in ballistic missile defense. The AF of this paper estimates the changing clutter spectrum parametrically and employs this parametric estimate to adaptively notch out the clutter returns. Conceptually the AF consists of a variable whitening filter (WF), followed by a variable, modified matched filter (MMF). The manner in which these filters are varied depends on the parametric spectral estimate of clutter determined from the data. This is shown in Figure 1. Note that this structure is different from the "estimator-subtractor" structure of sidelobe cancellers [1] or the least-mean-square (LMS) adaptive filter [2]. The present adaptive filter is an attractive alternative to the feedback structures of Applebaum [1] and Widrow [2]. The absence of feedback loops assures a shorter transient response for the filter, and the use of a Kalman filter for a spectral analysis provides a very rapid estimate of the clutter spectrum, thereby providing quick convergence of the WF and MMF to "match" the clutter environment. This latter feature is very attractive in rapidly changing clutter environments like tank break-up during reentry of ballistic objects.

The operation of the adaptive filter may be roughly explained as follows. The spectral analyzer estimates the clutter spectrum parametrically. Then the parameters are used to automatically design the whitening filter, or equivalently place notches at the clutter frequencies. Because the target signal of interest is modified in passing through the WF, the MMF is matched to the new signal. Thus, as the clutter spectrum is changing, the AF tries to maximize the signal to interference ratio adaptively. The technical approach in obtaining the AF structure is presented in Section 2.

This new adaptive filter has been successfully tested with synthetic data as well as with clutter from tank break-up in an actual ballistic missile test, as recorded at the Kwajalein Missile Range (KMR). Clutter suppression on the order of 20-30 dB better than the corresponding matched filter has been achieved with these data. These results are presented in Section 3 along with an analysis of the adaptive filter performance with a simulated clutter source (a Markov process). Some key points in the adaptive filter are discussed in Section 4. The paper concludes with a brief summary.

## 2. Technical Approach

We consider the detection of a target signal  $S(t)$  in clutter  $C(t)$  and white noise  $N(t)$  as a preliminary step in deriving the adaptive filter structure. This can be posed as a hypothesis testing problem. Under hypothesis  $H_1$  the target is present and the return signal  $R(t)$  can be written as

$$H_1: R(t) = S(t) + C(t) + N(t)$$

Under hypothesis  $H_0$  the target is not present and the return signal can be written as

$$H_0: R(t) = C(t) + N(t)$$

The objective is to be able to detect the target in a strong clutter environment, and minimize the false alarms due to clutter.

The first assumption in solving the problem is that clutter is a correlated noise process. In accordance with this assumption, let the autocorrelation function of  $C(t)$  be

$$\phi_C(t_1, t_2) = E [C(t_1) C(t_2)]$$

where  $E$  is the expected value operator. Let the thermal noise be a stationary process with a power spectral density of  $N_0/2$ . In this case, the optimum solution [3] can be obtained as shown in Figure 2.

The whitening filter (WF) decorrelates the interference  $C(t) + N(t)$  to a thermal noise like signal  $N_1(t)$ . In the process of doing so, the signal  $S(t)$ , if present, is modified to  $S_1(t)$ . The output of the WF is given by:

$$H_1: R_1(t) = S_1(t) + N_1(t)$$

$$H_0: R_1(t) = N_1(t)$$

This corresponds to the classical problem of detecting the known signal  $S_1(t)$  in the white noise  $N_1(t)$ . The solution consists of passing  $R_1(t)$  through a filter matched to the signal  $S_1(t)$ . This filter is called the modified matched filter (MMF). In other words the problem of detecting a signal in colored noise has been converted to a standard form. The mathematical details may be found in [3].

The impulse response of the whitening filter is given by the solution of an integral equation:

$$\int_{T_i}^{T_f} \phi_I(x, z) \int_{T_i}^{T_f} h(u, v) h(u, x) du dx = \delta(z-v) \quad T_i \leq z, v \leq T_f \quad (1)$$

where  $\phi_I(x, z) = \phi_c(x, z) + N_o/2 \delta(x-z)$

is the autocorrelation of the interference. Note that in the integral equation, the quantity inside the parenthesis is the inverse kernel of the correlation function of the interference. Naturally, the whitening filter impulse response  $h(t, \tau)$  is the solution to the integral equation with this inverse kernel. The above integral equation can be derived by assuming that the clutter process is nonstationary. This leads to a time varying whitening filter. In practice, clutter is not stationary, and the interval of observation is only finite. Therefore, the true optimum solution may be expected to involve a time varying whitening filter and MMF. However, if a short-time clutter spectrum is defined, it can be used to obtain the whitening filter. As the clutter spectrum varies with time, the impulse responses of the WF and MMF will vary insuring real-time adaptation to the changing environment.

In general the clutter correlation function  $\phi_c(t_1, t_2)$  is not known and must be estimated. Or, equivalently, the short term spectrum  $\phi_c(f)$  must be estimated, as in this study. Over a short period of time the clutter is modeled as a stationary autoregressive (AR) process; the parameters of the AR process are estimated; and, in turn, these estimates are used to automatically design the whitening filter and MMF. Because the estimates depend on the data, they change with time resulting in a time varying receiver structure.

Specifically, the clutter samples are modelled as an M-parameter autoregressive (AR) process. That is

$$C(k) = \sum_{i=1}^M \alpha_i C(k-i) + e(k) \quad (2)$$

where the  $\alpha_i$ 's are some constants,  $e(k)$  is a white noise process with unit variance and  $C(k-i)$  is the clutter sample lagging by  $i$  samples from the present sample. It is simple to show in this case that the power spectrum of the clutter process is given by

$$\phi_c(f) = \left| \frac{1}{1 - \sum_{i=1}^M \alpha_i \exp(-j2\pi f T)} \right|^2 \quad (3)$$

where  $T$  is the sampling interval.

Let  $\hat{\alpha}_i$  be the estimate of  $\alpha_i$  based on the observation of the clutter samples. Then, an estimate of the clutter spectrum is given by

$$\hat{\phi}_c(f) = \left| \frac{1}{1 - \sum_{i=1}^M \hat{\alpha}_i \exp(-j2\pi f T)} \right|^2 \quad (4)$$

In this case, the whitening filter (assuming a large clutter to noise ratio) is given by

$$H_w(f) = 1 - \sum_{i=1}^M \hat{\alpha}_i \exp(-j2\pi f T) \quad (5)$$

Note that the above whitening filter can be realized as a tapped delay line with tap weights given by 1 and  $-\hat{\alpha}_i$ , and has a simple structure which can be easily changed as the  $\hat{\alpha}$ 's vary. This was one of the motivating factors in modeling clutter as an AR process. Other factors include the noted success with which AR models have been applied to speech signals, which are also non-stationary processes.

Many different algorithms exist to estimate the parameters  $\alpha_i$ , and invariably involved time averages of correlation functions, and matrix inversions [4, 5, 6]. In this paper, a Kalman filtering technique is employed to estimate the WF parameters. The advantages are that no matrix inversion is involved in this case, and the estimates are obtained sequentially in real time.

To utilize Kalman filtering techniques, the spectral parameter estimation problem must be first converted to the state variable form. Accordingly the AR process parameters  $\alpha_1, \alpha_2, \dots, \alpha_M$  are defined as state variables and modeled as constants:

$$\underline{\alpha}(k+1) = \underline{\alpha}(k) \quad M \times 1 \text{ vector} \quad (6)$$

That is, the state transition matrix is an identity matrix. The process noise term, usually present in Kalman filtering is absent in this case. The observation is a scalar variable  $C(k+1)$  given by a time varying linear combination of the state variables  $\underline{\alpha}(k)$  plus the white noise term  $e(k+1)$ . This can be written compactly in a vector notion as

$$C(k+1) = \underline{H}(k+1) \underline{\alpha}(k+1) + e(k+1) \quad (7)$$

$$\text{where } \underline{H}(k+1) = [C(k) \ C(k-1) \ \dots \ C(k-M+1)] \quad (8)$$

Let the variance of the white noise component be denoted by  $G^2$ . By applying standard discrete Kalman filtering algorithm [7] to the message and observation models given by Eqs. (6) and (7), the following equations are obtained:

$$\hat{\underline{\alpha}}(k+1) = \hat{\underline{\alpha}}(k) + \underline{K}(k+1) [C(k+1) - \underline{H}(k+1) \hat{\underline{\alpha}}(k)] \quad (9)$$

$$\begin{aligned} \underline{V}_{\alpha}(k+1) &= [I - \underline{K}(k+1) \underline{H}(k+1)] \underline{V}_{\alpha}(k) \\ &\text{- covariance update equation} \end{aligned} \quad (10)$$

$$\begin{aligned} \underline{K}(k+1) &= \frac{1}{[G^2 + \underline{H}(k+1) \underline{V}_{\alpha}(k) \underline{H}^T(k+1)]} \\ &\text{- Gain Vector Computation} \end{aligned} \quad (11)$$

In the above equations, the superscript  $T$  indicates conjugate transponse. Note that  $\underline{H}(k+1) \hat{\underline{\alpha}}(k)$  is the one step prediction of the present clutter observation sample  $C(k+1)$ . The term  $[C(k+1) - \underline{H}(k+1) \hat{\underline{\alpha}}(k)]$  can be denoted by  $v(k+1)$  and is known as residual or innovation process. It is basically a sequence of "new information" and is a white process [8]. Note that the algorithm does not need to invert any matrix. This is because the observation  $C(k+1)$  is a scalar quantity.

The effect of using  $M$  parameters for the AR process is to interpolate the clutter spectrum with the frequency response of  $M$  poles in the complex frequency plane. The number chosen for  $M$  is generally a compromise between good spectral estimation and desirability of short transient response. It is also worthwhile noting that the AR process model can represent discrete target interference as well. Therefore, the method can also be used for reducing interference from other targets in a multiple target environment.

Certain drawbacks exist in the Kalman filter algorithm for clutter spectral estimation. If the clutter samples are present only up to a certain time, (say  $N$  samples), beyond which only noise is present, the estimates do not drop off to a small value, and are preserved with a memory typical of recursive techniques. In order to reduce this memory, a number of solutions have been proposed. One of the more attractive schemes is to have an

exponentially fading memory for the Kalman spectral estimator. This is implemented by artificially increasing the covariance  $V_{\alpha}(k)$  at each stage by a certain percentage [9]. This method has been fairly effective in the present study. Another method, called the limited memory filter [10], is a little more complicated, and periodically erases the memory of the estimator completely. In many cases, it is not clear how this periodicity can be chosen a priori. On the other hand, the exponentially fading memory length can be chosen such that the covariance is not changed by more than 20% each time. This increase reflects the uncertainty in the estimates, and is a divergence prevention technique. In such a case, the results are only near optimum.

### 3. Results

In this section, we present a number of examples illustrating the studies conducted with the proposed adaptive filter. In the first part of this section we derive the structure of the adaptive filter for the case of an N-pulse weighted burst waveform. This structure is used in the examples that follow. In the second part of this section we compare the signal-to-clutter (S/C) and signal-to-interference (S/I) ratios theoretically obtainable by an optimum interference rejection filter (the interference statistics exactly known), with that achieved by the adaptive filter for the case of a Markov random process as an interference source.

In the third part of the section we demonstrate the ability of the adaptive filter to place nulls in its frequency response corresponding to the interference frequencies. The final part of this section presents the application of the adaptive filter to suppress the tank break-up clutter of a re-entry booster in ballistic missile defense (BMD) application.

#### 3.1 Adaptive Filter Structure For An N-Pulse Burst

Let the transmit weights of an N-pulse burst waveform be  $\{a_1, a_2, \dots, a_N\}$  which may be complex in general. Then the sampled return from a desired target of unit strength, zero range, and doppler velocity  $\omega_d$  rad/sec will be

$$r(k) = a_k \exp(j\omega_d kT) \quad k = 1, 2, \dots, N \quad (12)$$

The matched filter to process this return will be a tapped delay line with weights  $\{b_k\}$  where

$$b_k = a_{N-k+1} \exp(-j\omega_d kT) \quad k = 1, 2, \dots, N \quad (13)$$

Without loss of generality we can set  $\omega_d=0$  in all our discussions, in which case the exponential in Eq. (13) can be set to unity. Let the spectral parameter estimates corresponding to the interference be  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_M$  at any given time sample. Then the impulse response of the whitening filter is given by  $\{1, -\hat{\alpha}_1, -\hat{\alpha}_2, \dots, -\hat{\alpha}_M\}$  and can be implemented as a tapped delay line. The return signal from the desired target is passed through the

whitening filter along with the interference. In this process, the target signal is modified and corresponds to the sequence  $\{c_k\}$ ,  $k=1, 2, \dots, M+N-1$  given by convolving  $\{a_k\}$  with  $\{1, -\hat{\alpha}_1, \dots, -\hat{\alpha}_M\}$ . In order to maximize the signal-to-noise ratio at the output of the processor, the filter following the whitening filter must be matched to this modified signal return  $\{c_k\}$ . Thus the modified matched filter (MMF) has an impulse response  $\{d_k\}$ , and is given by

$$d_k = c_{M+N-k}^*, k=1, 2, \dots, M+N-1 \quad (14)$$

The MMF can be implemented as a tapped delay line with weights  $d_k$ . The output of the MMF can now be processed further for detection processing.

The receiver structure for processing the N-pulse burst data is shown in Figure 3. This structure is basically used in our subsequent examples and discussions. Note that in practice we can combine the WF and MMF to implement a single tapped delay line filter with weights  $g_k$ ,  $k=1, 2, \dots, N+2M-1$  where  $\{g_k\}$  is the result of convolving  $\{1, \hat{\alpha}_1, -\hat{\alpha}_2, \dots, -\hat{\alpha}_M\}$  with the sequence  $\{d_k\}$ .

### 3.2 S/C and S/I Performance of the Adaptive Filter

Consider that the clutter is a stationary, first-order Gauss Markow process generated by

$$c_{k+1} = \theta c_k + w_k$$

where  $c_k$  is the clutter sample interference,  $|\theta| < 1$  and  $w_k$  is a white Gaussian sequence. Assuming zero mean for the processes involved, the covariance matrix of the interference for N samples is given by the  $N \times N$  Toeplitz matrix

$$\phi_c = \begin{bmatrix} 1 & \theta & \theta^2 & \dots & \theta^{N-1} \\ 0 & 1 & \theta & \dots & \theta^{N-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \theta^{N-1} & \theta^{N-2} & \ddots & \ddots & 1 \end{bmatrix}$$

Let the N-sample signal return be denoted by the vector  $\underline{s}$ . Let the return vector  $\underline{R}$  denote the sum of signal return, clutter, and thermal noise

$$\underline{R} = \underline{s} + \underline{c} + \underline{v}$$

where

$\underline{c}$  is an N vector of clutter samples

and  $\underline{v}$  is an N vector of thermal noise samples

Let the vector  $\underline{w}_{opt}$  denote the weights of the optimum tapped delay line

filter. Then the S/C and S/I ratios are given by

$$S/C = \frac{|W_{opt}^{\dagger} S|^2}{|W_{opt}^{\dagger} \phi_c W_{opt}|}$$

and

$$S/I = \frac{|W_{opt}^{\dagger} S|^2}{|W_{opt}^{\dagger} \phi_I W_{opt}|}$$

where  $\phi_I$  refers to the interference covariance matrix, the sum of the clutter and thermal noise covariance matrices  $\phi_c$  and  $\phi_v$ . For the case of the adaptive filter, we can obtain S/C and S/I with  $W_{opt}$  replaced by the adaptive filter weights  $W_{AF}$ .

In order to evaluate S/C and S/I ratios for the adaptive filter for various values of  $\theta$  a monte carlo simulation was performed. These ratios were computed at various stages of adaptation. To compare the performance of the adaptive filter with the optimum filter on the same basis, the noise gain of each filter was normalized to unity. The results of the simulation study are presented in Figure 4.

### 3.3 Adaptive Filter Frequency Response

The adaptive filter of this paper is derived essentially from a frequency domain viewpoint. The filter automatically adjusts its weight to place null in its frequency response corresponding to the significant frequency components of the clutter spectrum. We demonstrate this ability of the adaptive filter in this section.

Clutter data was simulated by means of two closely spaced discrete frequency scatterers. The clutter-to-noise ratio was 37 dB. Figure 5 presents the adaptive filter frequency response at the end of 15 adaptation samples. The clutter spectrum is also shown in the figure for comparison. Note that the null in the adaptive filter is down -100 dB from the zero frequency gain. The number of spectral parameters used in this case was 4.

### 3.4 BMD Applications

In a BMD environment, the clutter due to tank break-up (TBU) during re-entry into the atmosphere reduces the visibility of a threat (re-entry vehicle), and also increases the data processor load because of a large number of false alarms. In this case, the transmitted waveform is a coherent burst of 16 pulses with 50 dB Chebyshev weighting. The adaptive filter structure derived in the first part of this section was applied to this problem using data from an actual ballistic missile test as recorded by a radar at KMR. The data consisted of only the tank break-up clutter. The results are for the outputs of the matched filter (MF as well as the adaptive

filter for purpose of comparison. Figure 6 shows the MF and AF outputs for TBU data at a low altitude. This is a relatively low strength clutter region. From the figure, a clutter suppression on the order of 10 dB better than the matched filter is easily noted. Figure 7 shows the MF and AF outputs corresponding to TBU data from the heavy clutter region (high altitude). The spikes are ambiguities of a large tank fragment at the expected target doppler velocity, and is processed as a tank return, with no suppression. The additional clutter which is seen between any two spikes, is reduced by the adaptive filter significantly. Around the region of expected target location, the suppression is on the order of 20-30 dB more than the matched filter.

Many sets of these TBU data have been successfully processed with the adaptive filter demonstrating its clutter suppression capability in real world applications.

#### 4. Discussion

The adaptive filter presented in this paper is an open loop technique unlike the sidelobe canceller [1] or Widrow's adaptive filter [2]. This generally reduces the transient response duration. The convergence of the adaptive filter to the optimum filter depends on the convergence of the spectral estimator. In all our studies, this convergence has not been a problem. In particular in BMD applications where only 16 samples are available for processing, the conventional feedback techniques appear to be inadequate, while the new adaptive filter has performed satisfactorily. This seems to be primarily due to the rapid identification of clutter spectral parameters by the Kalman filter technique.

Although the results are generally satisfactory, some difficulties are encountered in using the new adaptive filter. First, a general theoretical analysis seems to be very difficult because of the time varying nature of the filter. This is partially solved in part 2 of the previous section where the Markov interference effects were analyzed experimentally. Another tool used often is the frequency response for particular experiments. This was done in part 3 of the previous section. Yet further analysis is necessary in relating the Kalman filter convergence, and the state error variances to the convergence of the adaptive filter. Also important is the study of the effects of divergence prevention techniques used in the Kalman filter. The key difficulty in developing the above analysis is that the variances in the Kalman filter are not pre-computable.

Secondly, note that the adaptive filter has an impulse response longer than the signal duration. This difficulty can be overcome by properly truncating the adaptive filter impulse response to signal duration. We can explain the increased duration of the impulse response as follows. Note that the theoretical impulse response of the whitening filter is given by Eq. (1).

This requires that the whitening filter be truly time varying for finite observation intervals. In our solution, we are replacing this time varying response by an estimate of the WF impulse response based on assumptions of an infinite observation interval and stationarity. This estimate, however, varies with time resulting in a time varying filter impulse response. But the time varying nature of the adaptive filter impulse response must be distinguished from that obtained by solving Eq. (1). In this sense the adaptive filter is only suboptimum, but the present implementation appears to be the only effective practical solution. It is worth noting that the conventional adaptive filters also have a similar difficulty in processing finite observation data. In other words, the time varying nature of the conventional adaptive filters is due to the changing estimates of the truly optimum weights, and does not represent the true time varying nature implicit in Eq. (1).

There is also another aspect to the time varying nature of all adaptive filters. If the interference environment is changing, these filters adjust their weight automatically to cancel out the interference. This time variation of the weights of the filter due to the changing environment is relatively slow compared to the sample-to-sample time variations due to random fluctuations.

In spite of the above mentioned shortcomings, the new adaptive filter has performed satisfactorily in practical applications, such as in Ballistic Missile Defense clutter suppression.

### 5. Conclusion

In this paper we presented a new open loop approach to adaptive filtering. By modeling the interference as an autoregressive process we could use a parametric spectral estimation technique directly yielding the whitening filter parameters. In particular, the Kalman filter was used to effectively identify the parameters. The clutter suppression ability of the new adaptive filter was demonstrated with applications to synthetic data and a ballistic missile defense problem.

A future paper will present the additional investigation currently in progress. These include further theoretical investigations, simple spectral estimation algorithms using stochastic approximation, and applications to spread spectrum communication, air-to-air seekers and airborne MTI.

### 5. References

1. Applebaum, S. O., 1976, Adaptive Arrays, IEEE Trans. Antennas and Propagation, Vol. AP-24, pp. 585-589.
2. Widrow, B., 1971, Adaptive Filters, in Aspects of Network and System Theory, Ed. by R. E. Kalman and N. DeClaris, Holt, Reinhart and Winston, Ch. 5.
3. Van Trees, H. L., 1968, Detection, Estimation and Modulation: Part I, John Wiley, Ch. 4.

4. Levinson, N., 1947, The Wiener RMS Error Criterion in Filter Design and Prediction, J. Math, Phys., Vol. xxv, pp. 261-278.
5. Srinath, M. D. and Viswanathan, M. M., 1975, Sequential Algorithm for Identification of Parameters of an Autoregressive Process, IEEE Trans. Automatic Control, Vol. AC.20, pp. 542-546.
6. Makhoul, J., 1975, Linear Prediction: A Tutorial Review, Proc. IEEE, Vol. 63, pp. 561-580.
7. Sage, A. P. and Melsa, J. L., 1971, Estimation Theory with Applications to Communication and Control, McGraw-Hill, Ch. 7.
8. Kailath, T., 1968, An Innovation Approach to Least Squares Estimation-Part I: Linear Filtering in Additive White Noise, Vol. AC-13, pp. 646-654.
9. Zarn, T. J., and Zaborsky, J., 1970, A Practical Nondiverging Filter, AIAA Journal, Vol. 8, pp. 1127-1133.
10. Jazwinsky, A. H., 1968, Limited Memory Optimal Filtering, IEEE Trans. Automatic Control, Vol. AC-13, pp. 558-563.

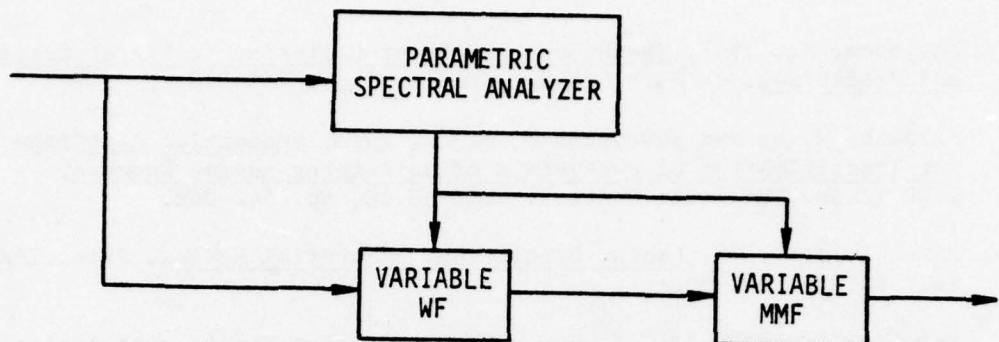


FIGURE 1: Conceptual Structure of the Adaptive Filter

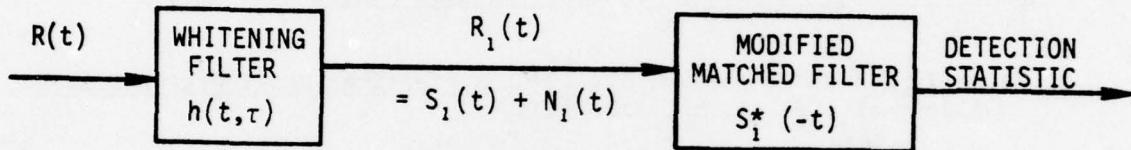


FIGURE 2: Optimum Receiver for Clutter Environment

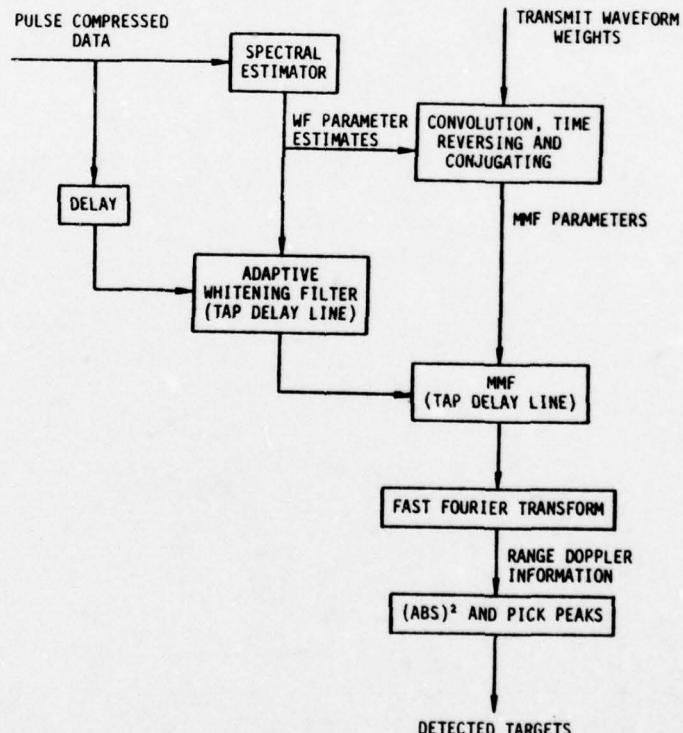


FIGURE 3: Adaptive Filter Structure to Process Coherent Burst of  $N$  Weighted Pulses

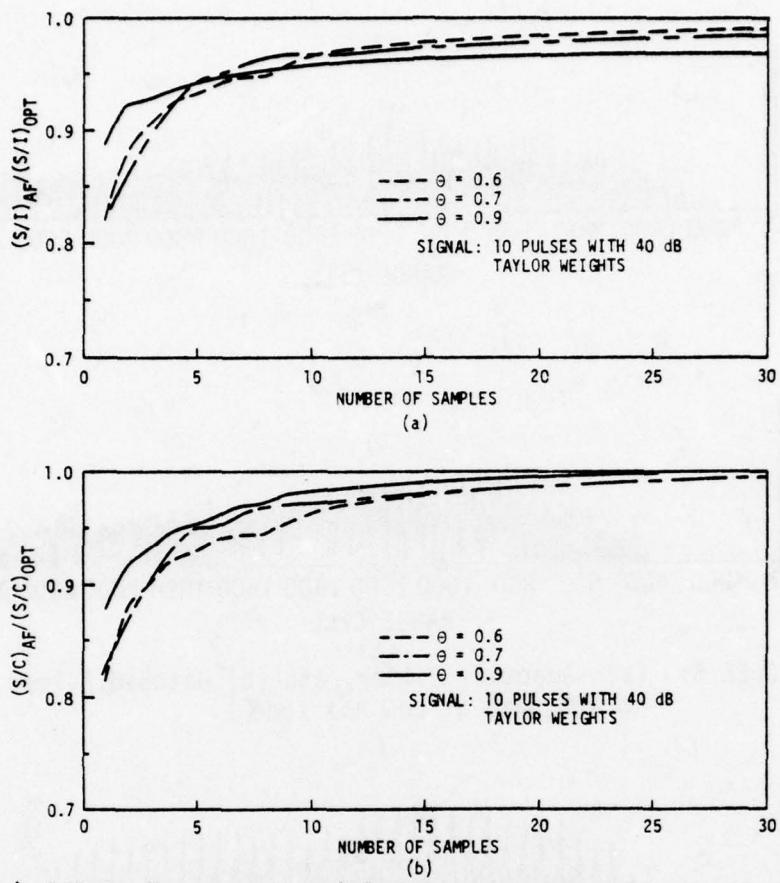


Figure 4:: (a) S/I Performance and (b) S/C Performance of the AF Relative to the Optimum Filter. (Signal is frequency offset from the clutter source)

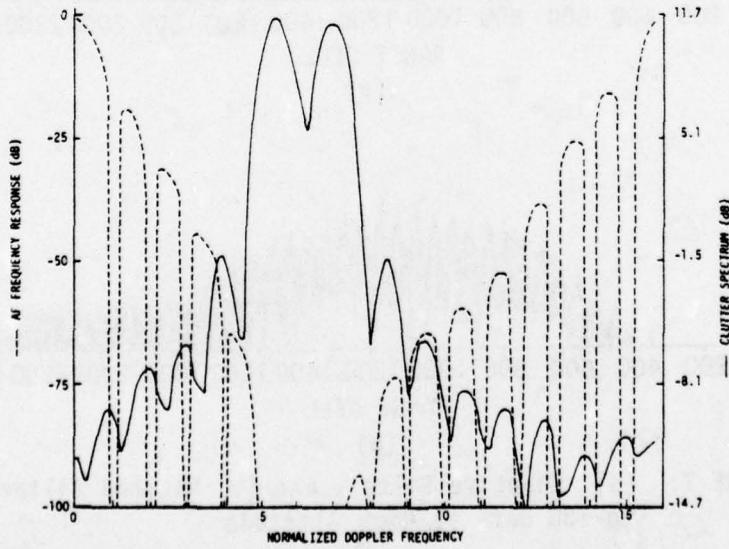
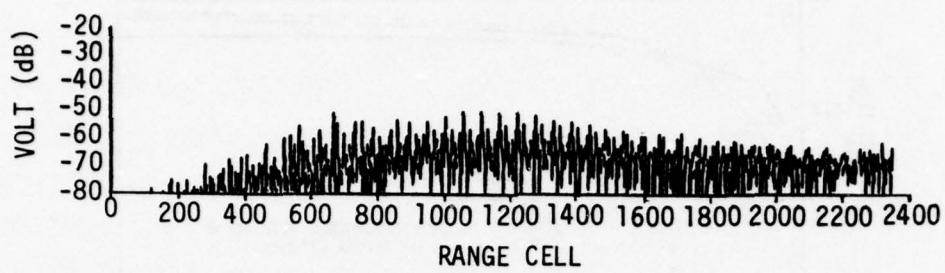
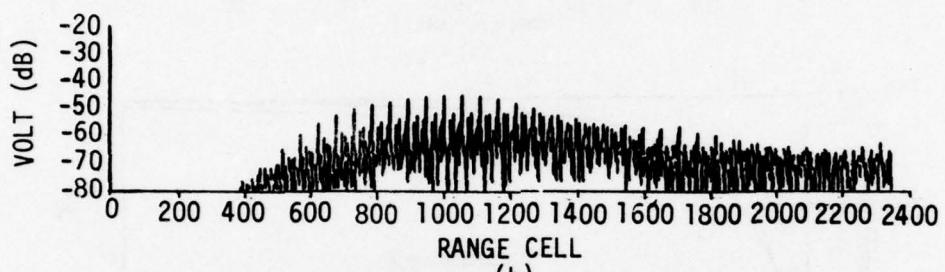


Figure 5: AF Frequency Response and Clutter Spectrum

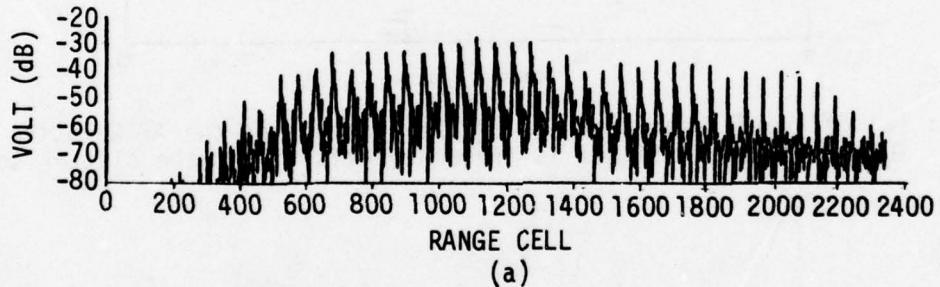


(a)

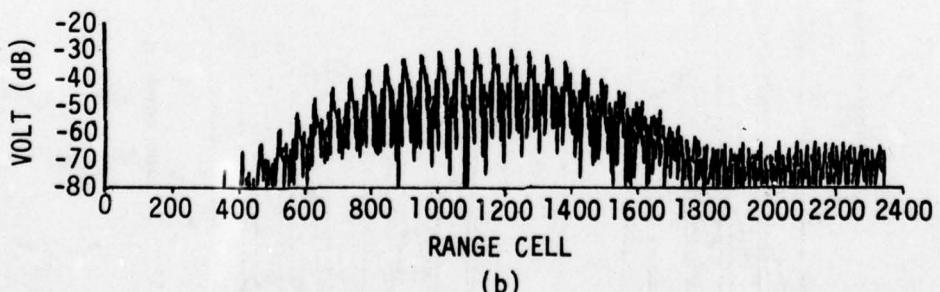


(b)

FIGURE 6: (a) Adaptive Filter, and (b) Matched Filter Response to TBU Data at Low Altitude



(a)



(b)

FIGURE 7: (a) Adaptive Filter, and (b) Matched Filter Response to TBU Data at High Altitude

DOPPLER SPECTRUM ESTIMATION  
FOR  
CONTINUOUSLY DISTRIBUTED RADAR TARGETS

GEORGE R. COOPER  
CLARE D. MCGILLEM

School of Electrical Engineering  
Purdue University  
W. Lafayette, IN 47906

Abstract

A method for estimating the doppler spectrum of targets having a continuous distribution in both range and velocity is discussed. This technique utilizes the random signal radar to achieve excellent range resolution and freedom from ambiguity, and the autoregressive spectral estimator to achieve excellent resolution in frequency. It is shown that the output of the wide-band correlator of the random signal radar has a spectral density that is a replica of the scatterer density function of the target in its velocity dimension. The autoregressive spectral estimator provides a computationally efficient method of estimating this spectral density with outstanding resolution characteristics.

Introduction

The class of targets considered here is made up of targets that are non-rigid and continuously distributed in range and velocity. This implies that a target can be modeled as a collection of scatterers that are in relative motion with respect to one another and are sufficiently close in range-doppler space that individual scatterers cannot be resolved. Examples of such targets are precipitation, weather formations, the ocean surface, foliage, chaff, etc. It is often of interest to know the distribution of scatterer velocities, particularly as a function of range. In meteorological measurements, this information makes it possible to distinguish vortex formations from non-vortex ones, or to measure wind shear as a function of altitude. Ocean surface measurements reveal wave heights and sea state and foliage measurements can determine the characteristics of certain classes of clutter and aid in intrusion detection [1], [2].

In most of the applications listed above, it is necessary to have excellent resolution in both range and velocity simultaneously. This requires that the radar signal have a large time-bandwidth product. The most widely used method of achieving large time-bandwidth products for radar signals is through the use of pulse compression by frequency modulation or pseudorandom binary phase coding of the transmitted signal. Such signals are periodic in nature and therefore exhibit ambiguous responses in both range and doppler.

Furthermore, they possess significant time and frequency sidelobes that result in erroneous responses for continuously distributed targets.

Many of the difficulties inherent in periodic waveforms can be overcome by using nondeterministic signals. The use of random signals has been thoroughly investigated both theoretically and experimentally and their utility demonstrated for many types of radar detection [1], [2], [3]. It is this type of signal that is considered in the present discussion.

#### Random Signal Radar Output

Only those aspects of the operation of the random signal radar that are essential to the present application are reviewed here. The transmitted signal consists of bursts of wideband random noise. Samples of this random signal are taken at the time of transmission, converted to binary form, and delayed in a multi-stage shift register. The received signal is also sampled and crosscorrelated with the delayed samples from the transmitted signal in a polarity-coincidence correlator. After filtering, the correlator output is proportional to the envelope of the crosscorrelation function of the transmitted signal and the received signal and provides a fully coherent detection of the received signal. A block diagram of such a system is shown in Figure 1.

The peak of the filtered output, as a function of correlator delay, occurs at  $\tau = 2R/c$  and corresponds to the round trip travel time to the target. The shape of the correlation function is determined by the spectral density of the transmitted signal and, hence, time sidelobes can be controlled readily with a simple filter or by shaping the sampling pulse. The range resolution is determined by the RF bandwidth of the transmitted signal and is closely approximated by

$$\Delta R = \frac{c}{2B} \quad (1)$$

where  $B$  is the bandwidth in Hz and  $c$  is the velocity of light.

When the target is moving, the envelope of the correlation function modulated by a sinusoid at the frequency of the doppler shift. Velocity resolution is given by

$$\Delta v = \frac{1}{2} \frac{\lambda}{c} W \quad (2)$$

where  $\lambda_c$  is the wavelength of the center frequency of the transmitted spectrum and  $W$  is the bandwidth (in Hz) of the correlator output filter centered at the doppler frequency being measured.

Since the transmitted signal is purely random, there is no ambiguity in range and the pulse repetition rate can always be made high enough to avoid any doppler ambiguities.

If the transmitted signal is  $x(t)$ , the return signal from the  $k$ th scattering point, having a radial velocity of  $v_k$ , is  $a_k x[t - \tau_k(t)]$  where  $a_k$  is

the relative magnitude of the return and

$$\tau_k(t) = \tau_k - \alpha_k t \quad (3)$$

in which  $\tau_k$  is round trip delay at  $t = 0$  and

$$\alpha_k = \frac{2 v_k}{c - v_k} \quad (4)$$

is the delay rate. If there are K such scattering points, the total signal into the correlator is

$$y(t) = \sum_{k=1}^K a_k x[(1+\alpha_k)t - \tau_k] + n(t) \quad (5)$$

where  $n(t)$  is the system noise.

The correlator also delays and time scales the reference signal so that it can be represented as

$$r(t) = x[(1+\alpha_r)t - \tau_r] \quad (6)$$

The received signal and the reference signal are multiplied together in the correlator and the product is filtered to remove all high-frequency components. Thus, the correlator output is an approximation to the expected value of  $[y(t)r(t)]$  and may be expressed as

$$E[y(t)r(t)] = R_{yr}(t) = \sum_{k=1}^K a_k R_x[(\alpha_r - \alpha_k)t - (\tau_r - \tau_k)] \quad (7)$$

where  $R_x(\cdot)$  is the autocorrelation function of  $x(t)$ . Since the transmitted signal is a bandpass function around some center frequency,  $f_c$ , its auto-correlation function will be of the form  $R_c(\tau) \cos \omega_c \tau$ . Thus, (7) becomes

$$R_{yr}(t) = \sum_{k=1}^K a_k R_c[(\alpha_r - \alpha_k)t - (\tau_r - \tau_k)] \cos \{\omega_c [(\alpha_r - \alpha_k)t - (\tau_r - \tau_k)]\} \quad (8)$$

Each term of this expression is oscillatory at a low frequency of  $(\alpha_r - \alpha_k)f_c$  and has an envelope that varies in accordance with  $R_c[(\alpha_r - \alpha_k)t - (\tau_r - \tau_k)]$ . Thus, the correlator output is the linear superposition of the coherent return from each of the scattering points.

If the scattering points are separated in range by an amount greater than  $\Delta R$ , as given by (1), or are separated in velocity by an amount greater than  $\Delta v$ , as given in (2), then they can be observed individually and the range and the velocity of each one measured. This capability of the radar may be a great advantage in identifying the characteristics of a rigid target.

However, in the case of continuously distributed targets, as considered here, the scatterers are not separated in either range or velocity by an amount sufficient to be resolvable. Hence, individual scattering points cannot be identified and the observer must be content with measuring the distribution of velocities in each range cell.

#### Model for Distributed Targets

In situations being considered here, it is convenient to define a scatterer density function,  $\rho(\alpha, \tau)$ , that is a measure of how the scattering points are distributed in both range and velocity. This density function is defined such that the quantity  $\rho(\alpha, \tau) \Delta\alpha \Delta\tau$  has the physical significance of being the average number of scattering points having a velocity (range rate) in the increment from  $\alpha$  to  $\alpha + \Delta\alpha$  and lying in the range increment corresponding to delays between  $\tau$  and  $\tau + \Delta\tau$ . Furthermore, the integral over all values of  $\alpha$  is proportional to the usual radar cross-section per unit volume.

It can be shown [4] that the expected value of the correlator output becomes

$$R_{yr}(t) = \iint \rho(\alpha, \tau) \overline{a(\tau)} R_x[(\alpha_r - \alpha)t - (\tau_r - \tau)] d\alpha d\tau \quad (9)$$

where  $\overline{a(\tau)}$  is the average attenuation coefficient of the returns from a range having a delay of  $\tau$  and may include the fourth power dependence on the range.

Although  $\rho(\alpha, \tau)$  is defined for distributed targets, it can be applied also to the discrete-target case by noting that in this case

$$\overline{a(\tau)} \rho(\alpha, \tau) = \sum_{k=1}^K a_k \delta(\alpha - \alpha_k) \delta(\tau - \tau_k) \quad (10)$$

Substituting (10) into (9) and carrying out the integration immediately yields the result given previously in (8).

When the scatterer density is smoothly distributed over a substantial range of  $\alpha$  and  $\tau$  values, so that  $\rho(\alpha, \tau)$  in (9) is changing slowly compared to the signal autocorrelation function, then (9) will yield a value that is very nearly zero. This is because  $R(\cdot)$  is a bandpass function and has zero net volume in  $\alpha$  and  $\tau$ . The physical meaning of this is that the correlator output, at any given time, consists of the linear superposition of many components of the form described by (8) with random frequencies and starting times, and that the instantaneous sum of these is just as likely to be negative as positive. As a consequence, the correlator output cannot be used to identify the parameters of any one scattering point as it could in the resolvable discrete target case.

However, in spite of the above difficulty, the correlator output does contain information regarding the scatterer density function,  $\rho(\alpha, \tau)$ . In particular, the time function  $R_{yr}(t)$  is a sample function from a random process having a well-defined spectral density that is related to the correlator delay and the velocity distribution of the scatterers at that delay.

This spectral density is given by [4]

$$S_R(f) = 2 P_T N_o + 2 P_c X_o + \frac{P_T X_o}{f_c} \overline{a^2(\tau_o)} \rho[\alpha_r + \frac{f-f_o}{f_c}, \tau_o] \quad (11)$$

where

$X_o$  = spectral density of the transmitted signal in bandwidth B

$P_T$  =  $2 X_o B$ , transmitted signal power

$N_o$  = spectral density of the receiver noise, assumed white

$P_c$  = total received signal power from all ranges

$f_o$  = a frequency off-set (greater than the largest doppler frequency) introduced in one channel of the receiver to avoid doppler foldover.

$\tau_o$  = delay of the reference signal

The total received signal power  $P_c$  is obtained from

$$P_c = \iint P_T \overline{a^2(\tau)} \rho(\alpha, \tau) d\alpha d\tau \quad (12)$$

and can be used to relate the scatterer density function to the radar cross-section per unit volume. If  $\eta(R)$  is the radar cross-section per unit volume as a function of range, then from the usual radar equation the total received signal power is

$$P_c = \frac{P_T \lambda_c^2}{\pi^2 \phi_b^2} \int \frac{\eta(R)}{R^2} dR \quad (13)$$

when it is assumed that the distributed target completely fills a circular beam of width  $\phi_b$  radians. Comparison of (12) and (13) indicates that

$$\eta(R) = \frac{2\pi R^2 \phi_b^2 \overline{a^2(2R/c)}}{c \lambda_c^2} \int \rho(\alpha, 2R/c) d\alpha \quad (14)$$

Thus, the radar cross-section per unit volume can be determined from the scatterer density function, which can in turn be measured by observing the spectral density of the correlator output.

### The Doppler Spectrum

The spectral density corresponding to (11) is shown in Figure 2 for a hypothetical situation. The flat portion of the spectrum, labeled  $S_{c+n}$ , is due to receiver noise and return signals from ranges other than those in the range cell having a delay of  $\tau_o$ . The peak in the spectrum comes from returns in the range cell of interest and has the shape of the scatterer density function. The objective of the spectrum measurement is to determine the magnitude,  $S_s$ , and shape of this peak. The accuracy with which this can be done is clearly a function of the ratio of  $S_s$  to  $S_{c+n}$  and the frequency resolution capability of the spectral estimation procedure. This aspect is discussed subsequently.

Letting  $v$  be any value of radial velocity, then it is possible to define a velocity density,  $\rho(v)$ , where  $\rho(v)\Delta v$  has the physical significance of representing the fraction of the total return from the range cell of interest that come from scattering points having a relative velocity in the interval from  $v$  to  $v + \Delta v$ . This velocity density is related to the spectral density by

$$\rho(v) = \frac{1}{A} [S_R \left( \frac{2v}{\lambda_c} + f_o \right) - S_{c+n}] \quad (15)$$

where  $A$  is the area under the peak of the spectral density and is defined by

$$A = \int_0^{\infty} [S_R(f) - S_{c+n}] df \quad (16)$$

It is clear from the above discussion that an estimate of the spectral density of the correlator output over a sufficient range of frequencies will lead directly to an estimate of the velocity density of the distributed target in the range cell of interest. If the range at which the measurement is made is changed, and if the measurements are repeated in a sufficient number of time intervals, then a complete history of the velocity density function as a function of range and time can be obtained.

### Accuracy Considerations

Before describing a technique for estimating the spectral density, it is appropriate to state some fundamental limits on the accuracy with which this spectrum can be estimated. If the estimate of  $S_R(f)$  is denoted by  $\hat{S}_R(f)$ , a convenient measure of the goodness of the estimate is its signal-to-noise ratio defined by

$$(S/N)_o = \frac{E^2[\hat{S}_R(f)]}{\text{Var}[\hat{S}_R(f)]} \quad (17)$$

where  $E[\hat{S}_R(f)]$  is the mean value of the estimate of any specified frequency and  $\text{Var}[\hat{S}_R(f)]$  is the variance of the estimate.

It has been shown [6] that this signal-to-noise ratio can be expressed in the form

$$(S/N)_o = 2[1 + 2WT] \left[ \frac{I}{I+1} \right]^2 \quad (18)$$

where  $W$  is the desired frequency resolution,  $T$  is the time interval over which the correlator output is observed, and

$$I = \frac{s_s}{s_{c+n}} \left( \frac{f_s}{2B} \right) \quad (19)$$

where  $f_s$  is the rate at which samples are taken from each transmitted pulse and  $B$  is the transmitted signal bandwidth.

Two things are immediately apparent; the quantity  $I$  must be larger than one if the last factor in (18) is not to have a significant effect, and if  $I \gg 1$  then  $(S/N)_o$  is determined almost entirely by the desired frequency resolution  $W$  and the observation time  $T$ . It is important to note that better frequency resolution (that is, smaller values of  $W$ ) leads to poorer signal-to-noise ratios.

As a means of indicating the magnitudes of some of these quantities, consider the numerical values in Table I.

Table I. Numerical Values Used For Illustration

$P_T = 1 \text{ W}$	$f_c = 9 \text{ GHz}$
$\phi_b = 0.018 \text{ radians } (1^\circ)$	$N_o = 1.644 \times 10^{-20} \quad (T_{\text{eff}} = 1200^\circ \text{K})$
$B = 75 \text{ MHz}$	$\eta = 10^{-7} \text{ m}^2/\text{m}^3$
$R_o = 1000 \text{ m}$	$W = 120 \text{ Hz}$
$T = 0.1 \text{ seconds}$	$f_s = 15 \text{ MHz}$

The radar cross-section per unit volume,  $\eta$ , appearing in the above table corresponds to rainfall at a rate of 1.25 mm/hr. For these parameters the quantity  $I$  turns out to be  $I = 3.27$  and the resulting signal-to-noise ratio is

$$(S/N)_o = 2 [1 + 2 \times 120 \times 0.1] \left[ \frac{3.27}{3.27+1} \right]^2$$

$$= 29.3 \quad \text{or} \quad 14.7 \text{ dB}$$

Although this value is not large, it is adequate for many measurements.

#### Doppler Spectrum Estimation

There are many different ways in which the spectral density of the correlator output can be measured. The most obvious method is the use of

filters and a square-law envelope detector as shown in Figure 3. This will estimate the spectral density at one frequency. In order to obtain a complete spectrum, the bandpass filter must either be tunable or a bank of bandpass filters must be used.

A more effective technique for this application, however, appears to be the use of the autoregressive spectral estimator. This technique has been widely discussed in the literature [7], [8] and an empirical comparison with other standard technique has been made [9]. The basic philosophy of this approach is to fit a finite-order autoregressive process (AR) to the observed spectrum in a minimum mean-squared error sense. This leads to a computationally efficient spectral estimate that is capable of excellent resolution.

If  $\{z_k\}$  are a sequence of equally spaced samples from a random process, then an  $L$ th order AR process is defined by the difference equation

$$z_k = \sum_{i=1}^L a_i z_{k-i} + n_k \quad (20)$$

where the  $n_k$  are samples of a stationary Gaussian white noise process having zero mean and a variance of  $\sigma_n^2$ .

In order to estimate the spectral density of the process from which the samples are taken, the following quantities need to be estimated:

- 1) Samples of the estimated autocorrelation are obtained from

$$\underline{r}(i) = \frac{1}{N-1} \sum_{j=0}^{N-i-1} z_j z_{j+i}, \quad i = 0, 1, \dots, L \quad (21)$$

after subtracting out the sample mean. When  $N$  is large, this computation can be done more efficiently through the use of the Fast Fourier Transform.

- 2) The AR coefficients,  $\{a_i\}$ , can be estimated from the  $\{\hat{r}(i)\}$  by solving the matrix equation

$$\hat{\underline{R}} \hat{\underline{a}}^T = \hat{\underline{r}}^T \quad (22)$$

where

$$\begin{aligned} \underline{r} &= [r(1), \dots, r(L)] \\ \hat{\underline{a}} &= [a_1, \dots, a_L] \\ \hat{\underline{R}} &= \begin{bmatrix} \hat{r}(0) & \hat{r}(1), \dots, \hat{r}(L-1) \\ \vdots & \\ r(L-1) & r(0) \end{bmatrix} \end{aligned}$$

This equation can be solved efficiently using the Levinson algorithm.

- 3) Estimate the white noise variance,  $\hat{\sigma}_n^2$ , from

$$\hat{\sigma}_n^2 = \hat{r}(0) - \sum_{i=1}^L \hat{a}_i \hat{r}(i) \quad (23)$$

The Lth-order AR estimate of the spectral density is then obtained from

$$\hat{S}_R(f) = \frac{\hat{\sigma}_n^2 \Delta T}{\left| 1 - \sum_{i=1}^L a_i e^{-j2\pi f i \Delta T} \right|^2}, \quad |f| \leq \frac{1}{2\Delta T} \quad (24)$$

where  $\Delta T$  is the sampling period.

Since the AR spectral estimator is nonlinear, analytical derivation of its statistical properties is very difficult. However, some asymptotic properties have been derived in the literature [11], [12]. These results indicate that the AR estimator has asymptotic statistical properties that are similar to the Fourier transform methods using a rectangular window.

An empirical investigation of the frequency resolution and accuracy properties of the AR spectral estimator has also been done [9]. This study indicates that a specified resolution for an AR estimator can be obtained with a much smaller number of lag values (i.e., L) than is possible with Fourier transform methods. Furthermore, the frequency-averaged variance of the AR spectral estimate is essentially the same as for Fourier transform methods using a rectangular window for a given ratio of L/N. However, since L can be smaller for the AR estimator for a given frequency resolution, the net result is that the frequency-averaged variance of the AR technique will be smaller than that of the Fourier transform technique when they are compared on the basis of equivalent resolving power.

#### Experimental Results

The techniques described above have been used with the random signal radar to estimate the doppler spectrum of rainfall. The characteristics of the radar are displayed in Table II.

Table II. Random Signal Radar Used for Estimating  
Doppler Spectrum of Rainfall

Frequency, $f_c$	8.94 GHz
Average transmitted power/pulse, $P_T$	26 mW
Pulse repetition rate	1 MHz
Sampling rate, $f_s$	1 MHz
Range resolution, $\Delta R$	4.5 m
Antenna	30" paraboloid
Range, $R_o$	100 m
Observation time, $T$	0.36 seconds

Figure 4 shows the estimated doppler spectrum computed from  $N = 3000$  samples ( $\Delta T = .12$  ms) for several different values of  $L$ . This figure shows that the spectral peak is fully resolved for  $L = 250$  [13].

#### Conclusions

The main point to be emphasized in this discussion is that the excellent range resolution characteristic of the random signal radar can be combined with the excellent frequency resolution characteristic of the autoregressive spectral estimator to obtain a system that is capable of measuring the complete doppler spectrum of small regions in a continuously distributed radar target. This combination of characteristics is very difficult to achieve with conventional pulse radar without encountering serious ambiguity problems. Thus, the approach described here provides a useful technique for a wide range of problems involving distributed targets, discrete targets embedded in distributed targets, moving clutter measurement or any situation in which there is relative motion of scatterers that cannot be resolved in either range or doppler.

#### References

1. Chadwick, R.B. and G.R. Cooper, "Measurement of Ocean Wave Heights with the Random Signal Radar," IEEE Trans. on Geoscience Electronics, Vol. GE9, No. 4, October 1971, pp. 216-221.
2. Cooper, G.R. and C.D. McGillem, "Complex Target Resolution with the Random Signal Radar," EASCON Record, 1972.
3. McGillem, C.D., G.R. Cooper, and W.B. Waltman, "Use of Wide-and Stochastic Signals for Measuring Range and Velocity," EASCON Record, 1969, pp. 305-311.
4. Chadwick, R.B. and G.R. Cooper, "Measurement of Distributed Targets with the Random Signal Radar," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-8, No. 6, November 1972, pp. 743-750.

5. McGillem, C.D., G.R. Cooper, and W.B. Waltman, "An Experimental Random Signal Radar," Proc. NEC, Vol. XXIII, 1967, pp. 409-412.
6. Chadwick, R.B., "Use of Random Signals to Study Distributed Radar Targets," Ph.D. Thesis, Purdue University, August 1970.
7. Parzen, E., "Statistical Spectral Analysis (Single Channel Case) in 1968," Stanford University Technical Report 11 on Contract Nonr-225(80), June 10, 1968.
8. Burg, J.P., "Maximum Entropy Spectral Analysis," paper presented at 37th Annual International SEG Meeting, Oklahoma City, Oklahoma, October 31, 1967.
9. Kaveh, M. and G.R. Cooper, "An Empirical Investigation of the Properties of the Autoregressive Spectral Estimator," IEEE Trans. on Information Theory, Vol. IT-22, No. 3, May 1976, pp. 313-323.
10. Levinson, N., "The Wiener RMS Error Criterion in Filter Design and Prediction," J. Math. Phys., Vol. 25, 1946, pp. 261-278.
11. Kromer, R.E., "Asymptotic Properties of the Autoregressive Spectral Estimator," Ph.D. Dissertation, Stanford University, December 1976.
12. Berk, K.N., "Consistent Autoregressive Spectral Estimates," Ann. Statist., Vol. 2, May 1974, pp. 489-502.
13. Kaveh, M., "High Resolution Velocity Estimation in the Presence of Extended Clutter," Ph.D. Thesis, Purdue University, December 1974.

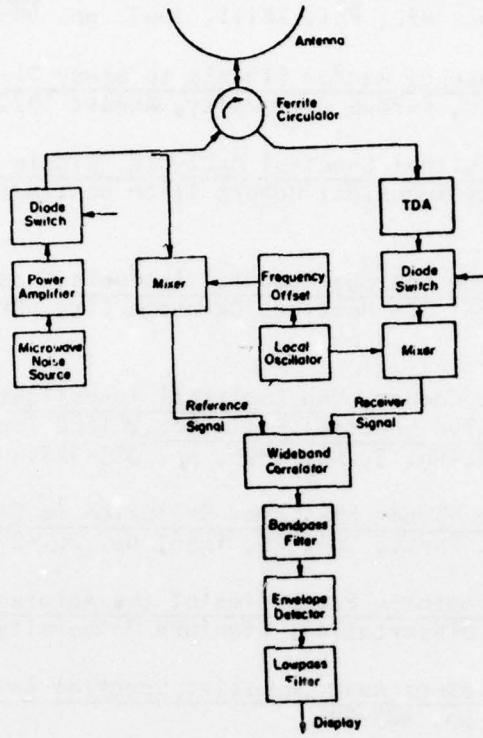


Figure 1. Block diagram of the random signal radar.

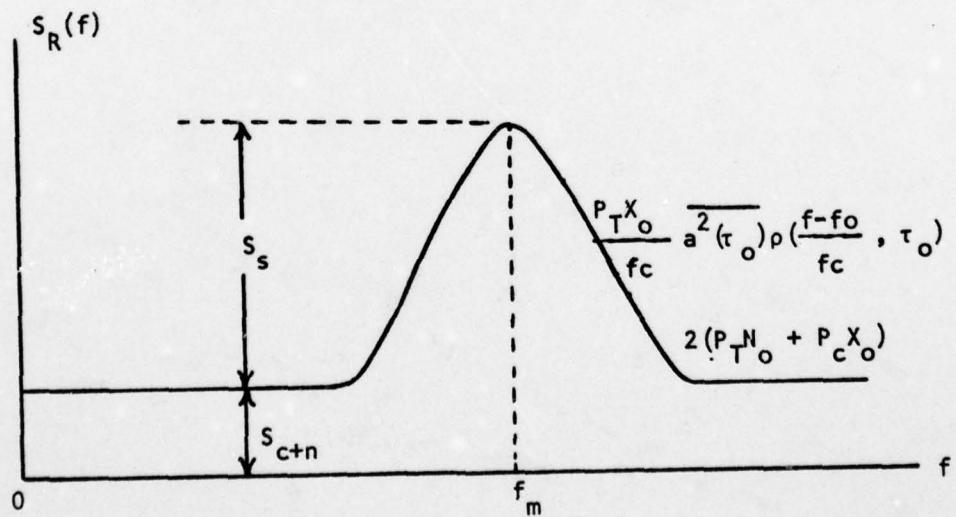


Figure 2. Spectral density of the correlator output.

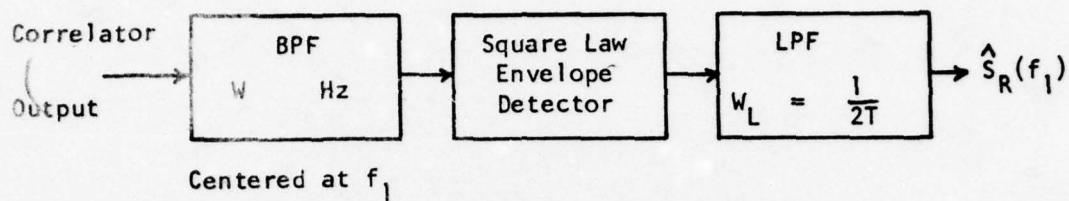


Figure 3. Filter method of estimating spectral density.

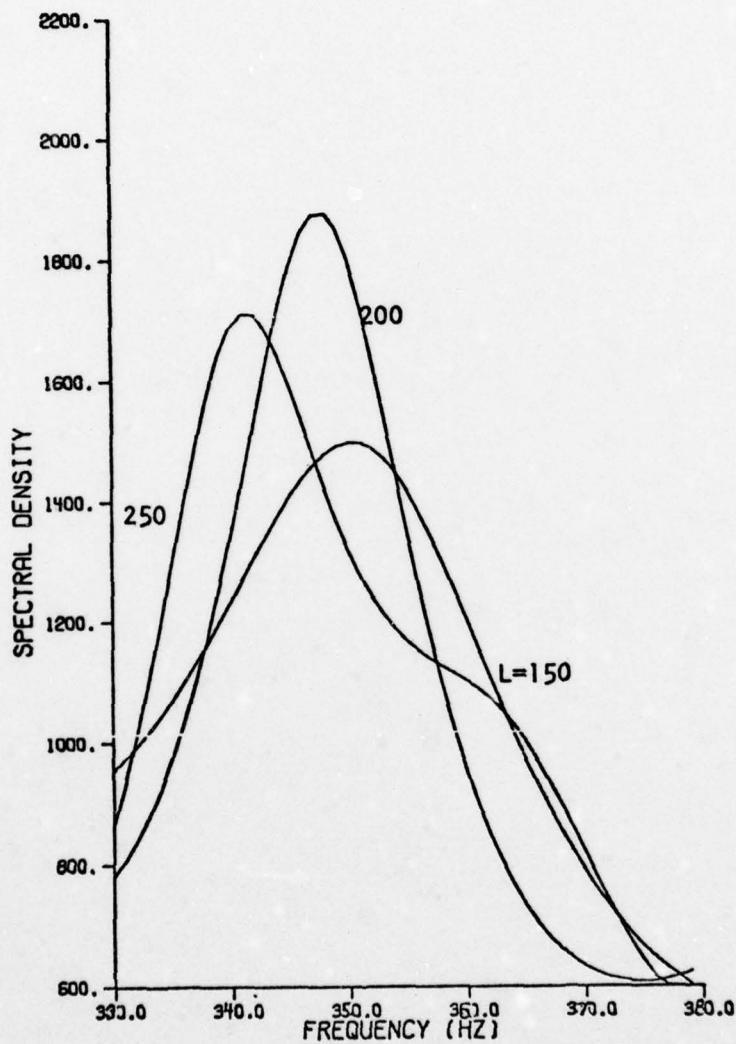


Figure 4. Doppler spectrum for rainfall with the antenna pointing 90° above the horizon [13].

INSTANTANEOUS FREQUENCY ESTIMATION  
FROM SAMPLED DATA

WILLIAM R. CARMICHAEL  
DR. RICHARD G. WILEY

Abstract

Estimating the instantaneous frequency of a frequency modulated signal in the presence of noise is simplified when the input signal-to-noise ratio (SNR) is greater than about 12 dB. The techniques described here use the zero crossings of the signal as a basis for estimating the instantaneous frequency of data provided by the Rome Air Development Center (RADC). The estimated rms error is 0.7%.

Introduction

Preliminary Analysis

For white noise and an input SNR of 12 dB or more, it is well known that the SNR out of an FM demodulator is given by

$$\text{SNR}_{\text{out}} = 3 \text{ SNR}_{\text{in}} \left( \frac{\Delta F}{F_m} \right)^2 \quad (1)$$

where:

$\Delta F$  = peak frequency deviation  
 $F_m$  = highest modulating frequency

For the signal data supplied by RADC, the input SNR is given as about 20 dB, and the input signal to noise plus interference ratio is about 17 dB. Equation 1 may be applied in this case since the interference is small.

RADC also specified that the spectrum of interest was from 130 to 270 Hz and that the rate of frequency drift was no greater than 190 Hz/s.

For a single sinusoidal modulation, the maximum rate of change of frequency is given by

$$\text{Maximum Slope} = \Delta F \cdot 2\pi F_m \quad (2)$$

The largest modulating frequency is, therefore,

$$\text{Maximum } F_m = \frac{(\text{Maximum Slope})}{2\pi \Delta F} \quad (3)$$

For a given slope, the maximum modulating frequency is largest when the peak deviation is the smallest. Examining the spectrum of the input (see Figure 1A) shows a total frequency deviation of at least 60 Hz. Then the minimum peak deviation is 30 Hz and the maximum modulating frequency is

$$\text{Maximum } F_m = \frac{190}{2\pi \cdot 30} \approx 1 \text{ Hz} \quad (4)$$

In this situation, the output SNR should be approximately

$$\text{SNR}_{\text{out}} = 17 \text{ dB} + 10 \log[3(30)^2] = 51 \text{ dB} \quad (5)$$

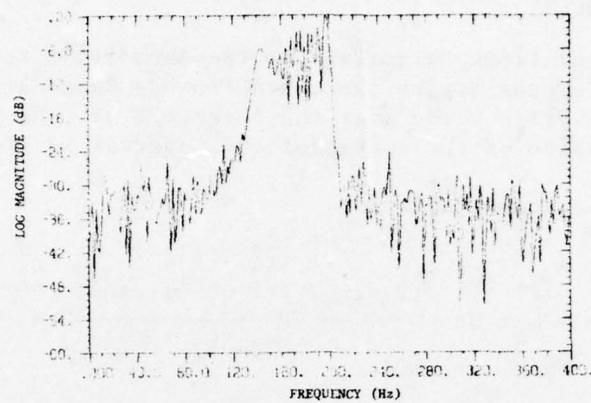
Therefore, an ordinary FM demodulator will provide good performance. Note, however, that the output bandwidth of such a system would be on the order of 1 Hz. Because the data sample available is only 0.64 s in duration, the substantial transient effects due to use of a 1 Hz filter would severely affect the estimates of the instantaneous frequency within the 0.64 s record.

#### Data Prefiltering

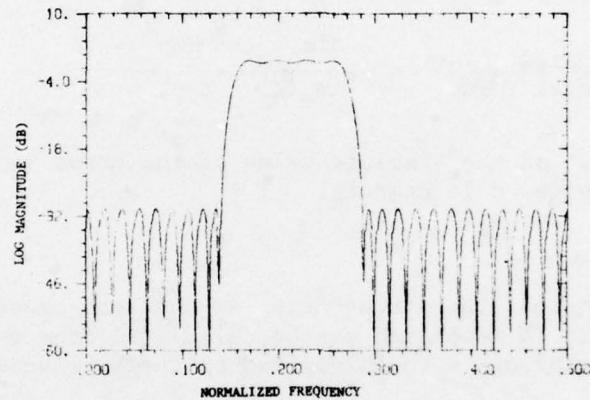
Figure 1A presents a plot of the spectrum of the raw input data. Since the data is narrowband, bandpass filtering was applied to improve the SNR. The filter of Figure 1B is a 63 point FIR filter designed with McClellan's Remez Exchange Algorithm [4]. The filter passband extends from 130 Hz to 210 Hz for a sampling rate of 800 Hz. Note, the upper band edge of the data was estimated to be 210 Hz; the value given in the problem statement was 270 Hz. An FIR filter was chosen because of its ideal phase characteristic. The peak in band amplitude ripple is -30 dB. In Figure 1C, the spectrum of the data after band-pass prefiltering is presented.

#### Zero Crossings and Frequency Estimation

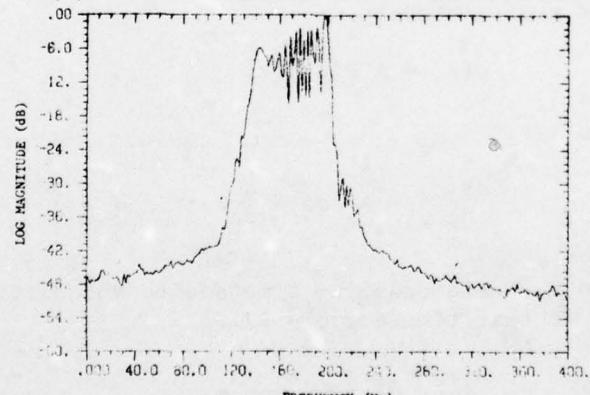
The average frequency of the signal (in the absence of noise) is given by the reciprocal of twice the time between two zero crossings. In this case, there will be many zero crossings per second compared to the bandwidth of the modulation. Assume that the modulating signal is bandlimited. Then, in view of the mean value theorem for continuous differentiable functions, the instantaneous frequency will be equal to the average frequency at some point between the zero crossings used to estimate the average frequency. In Appendix A, it is demonstrated that the instantaneous frequency at the midpoint of the interval is equal to the average frequency if the modulating signal is linear over that time. Since there are many zero crossings per cycle of the highest modulating frequency present, the error caused by assuming that the modulating signal is approximated by a straight line over a few zero crossing intervals will be small.



(A) RAW INPUT DATA SPECTRUM



(B) FIR BANDPASS FILTER CHARACTERISTIC



(C) PREFILTERED DATA SPECTRUM

Figure 1. Data Spectra Before and After Filtering

### Linear Approximation Error

Assume that the modulating signal is of the form  $\sin \omega_m t$ . Here we determine the error due to approximating the instantaneous value at the center of the time interval by the average value over the interval. If the interval is  $(t_1, t_2)$ , the actual value at the center of the interval is given by

$\sin \omega_m \left( \frac{t_2 + t_1}{2} \right)$ . The average value is

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \sin \omega_m t dt = \frac{2 \sin \omega_m \left( \frac{t_2 + t_1}{2} \right) \sin \omega_m \left( \frac{t_2 - t_1}{2} \right)}{\omega_m (t_2 - t_1)} \quad (6)$$

The ratio of the estimated value to the actual value of the function is

$$\frac{\text{Estimated Value}}{\text{Actual Value}} = \frac{2 \sin \omega_m \left( \frac{t_2 - t_1}{2} \right)}{\omega_m (t_2 - t_1)} \quad (7)$$

Figure 2 is a plot of the absolute value of the error in dB versus the averaging interval expressed in degrees.

### Error Due to Noise

In the high SNR cases, the perturbation of the zero crossings of a sinusoidal signal by noise will be minor and can be calculated from the slope of the sinusoid as it passes through zero multiplied by the amplitude of the noise at that instant.

Let the signal be represented by

$$s(t) = A \sin \theta(t) \quad (8)$$

Differentiating gives

$$\frac{ds(t)}{dt} = A \cos \theta \frac{d\theta}{dt} \quad (9)$$

Then the error in the zero crossing time due to an additive signal of magnitude N will be (since at that time  $\cos \theta = 1$ )

$$\Delta T = \frac{N}{A \frac{d\theta}{dt}} \quad (10)$$

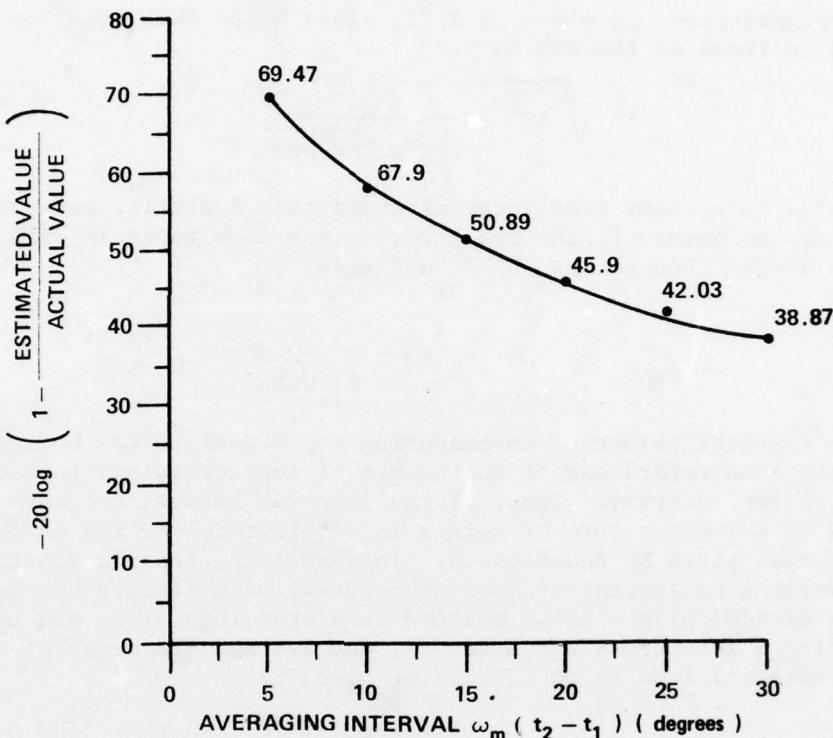


Figure 2. Linear Approximation Error

Since  $d\theta/dt$  is the instantaneous frequency multiplied by  $2\pi$ ,

$$\Delta T = \frac{N}{2\pi A F_i} \quad (11)$$

where:

- $\Delta T$  = error in zero crossing time due to noise  $N$ ,
- $N$  = instantaneous noise amplitude at zero crossing time,
- $A$  = signal peak amplitude, and
- $F_i$  = instantaneous frequency.

Because the noise can be considered a normally distributed random variable independent from one zero crossing to the next, the rms value of the error in the zero crossing time is simply

$$\sqrt{\Delta T^2} = \frac{\sqrt{N^2}}{2\pi A F_i} \quad (12)$$

The signal power is given by  $A^2/2$ . This means that Equation 12 can be expressed in terms of the SNR as

$$\sqrt{\overline{\Delta T^2}} = \frac{1}{2\pi F_i \sqrt{2 \text{SNR}}} \quad (13)$$

Finally, since any time interval would have a similar independent error at each end of the interval, the rms error in any such interval will be  $\sqrt{2}$  times the value above. Denote this by  $\sigma_T$  and then

$$\sigma_T = \frac{1}{2\pi F_i \sqrt{\text{SNR}}} \quad (14)$$

Note that the rms error in measuring any time interval between zero crossings is the same regardless of the number of zero crossings included between the ends of the interval. Thus, if the interval between two zero crossings separated by  $K-1$  other zero crossings is considered, the rms error in that interval is also given by Equation 14. In that time, the phase will have advanced by  $K\pi$ , thus the reciprocal of such an interval will measure the instantaneous frequency divided by  $K$ . If we measure zero crossings at  $t_1$  and  $t_2$  and count  $t_2$  as occurring  $K$  zero crossings after  $t_1$ , the average frequency of the signal over the interval from  $t_1$  to  $t_2$  will be given by

$$F = \frac{K}{2(t_2 - t_1)} \quad (15)$$

We might then designate the value in Equation 15 as the instantaneous frequency at the midpoint of the interval (subject to the linear approximation error already discussed). The error in this frequency estimate due to an error in estimating the interval  $(t_2 - t_1)$  is

$$\Delta F = \frac{-K \Delta T}{2(t_2 - t_1)^2} \quad (16)$$

Using Equation 14, the rms frequency error is expressed as

$$\sigma_F = \frac{K \sigma_T}{2(t_2 - t_1)^2} = \frac{K}{4\pi F_i \sqrt{\text{SNR}} (t_2 - t_1)^2} \quad (17)$$

Note that, approximately,  $F_i(t_2 - t_1) = K/2$ , so that

$$\frac{\sigma_F}{F_i} = \frac{1}{\pi K \sqrt{\text{SNR}}} \quad (18)$$

Hence, an improvement in the SNR at the output proportional to the square of the number of zero crossings separating  $t_2$  from  $t_1$  is expected, i.e.

$$\text{SNR}_{\text{out}} = \pi^2 K^2 (\text{SNR})_{\text{in}} \quad (19)$$

Since large  $K$  implies large time intervals, as  $K$  increases, more error due to approximating the instantaneous frequency at the midpoint of the interval by the average over the interval is introduced.

A typical instantaneous frequency in the data record is 170 Hz. If  $F_m = 1$  Hz, for  $\omega_m(t_2 - t_1) = 20^\circ$ ,  $(t_2 - t_1)$  will be approximately 0.056 s. In this case,  $K \approx 2 \times 170 \times (0.056) \approx 19$ . The output SNR will then be 52.5 dB according to Equation 19, and the output signal to approximation error will be 45.9 dB according to Figure 2. The combined SNR from both errors would be about 45.0 dB. This approach to estimating the instantaneous frequency would yield an accuracy of about

$$\frac{\Delta F}{F} = \frac{1}{10^{2.25}} = 0.56\%,$$

which seems adequate for most applications.

Furthermore, the approach is quite simple:

- Step 1 locate the zero crossings.
- Step 2 compute the reciprocal of twice the interval between those zero crossings separated by about 0.056 s.
- Step 3 multiply those estimates by the number of zero crossings separating the ends of each such interval.
- Step 4 assign this value as the instantaneous frequency at the center of the interval.

#### Final Interpolation

The results of the steps above are instantaneous frequency estimates for various times which are neither equally spaced nor located at the times for which estimates are desired. The estimates are available at intervals of about 0.028 s, and the desired outputs are located at multiples of 0.04 s. If the modulation is at a 1 Hz rate, we obtain a sample about every  $10^\circ$ . With samples spaced  $10^\circ$  apart, the sine function may be interpolated linearly to 0.4%. Since the error caused by this is slightly less than the error involved in obtaining the instantaneous frequency estimates at the interval midpoints, this is a satisfactory technique.

AD-A054 650      ROME AIR DEVELOPMENT CENTER GRIFFISS AFB N Y      F/G 14/2  
PROCEEDINGS OF THE RADC SPECTRUM ESTIMATION WORKSHOP HELD ON 24--ETC(U)  
MAY 78

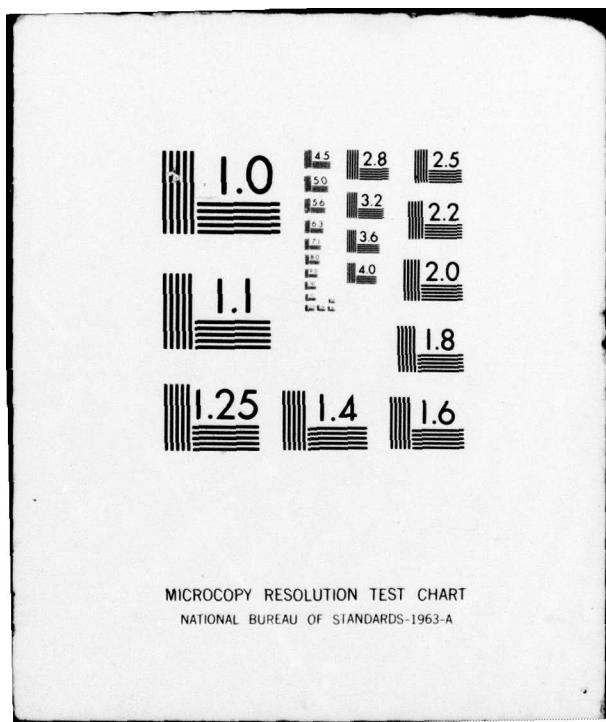
UNCLASSIFIED

NL

4 OF 4  
AD  
A054650



END  
DATE  
FILMED  
6 -78  
DDC



### Zero Location Procedure

The original data is sampled at 800 Hz. Since the frequency approaches 200 Hz, we have samples at 90° intervals in the worst case. It would be convenient to locate the zero crossings by linear interpolation whenever the data changes sign. However, this would introduce errors of up to 5% of the sampling interval of 0.00125 s, as shown by Figure 3. This would cause an error in the instantaneous frequency estimates of

$$\frac{\Delta F}{F_i} = \frac{K \Delta T}{2(t_2 - t_1)^2 F_i} = \frac{\Delta T}{(t_2 - t_1)}$$

$$\frac{\Delta F}{F_i} = \frac{\Delta T}{T_2 - T_1} = \frac{(0.00125)(0.05)}{0.056} = 0.11\%$$

While this is sufficiently small, we chose to increase the sampling rate by a factor of five using a finite impulse response low pass filter [5]. This allowed for samples at 4000 Hz. In the worst case of 200 Hz for the signal, there are then 20 samples per cycle (18°) and the worst case error was reduced to 0.15% of the sampling interval or  $\Delta F/F_i = 7.7 \times 10^{-5}$  due to the use of linear interpolation to initially locate the zeros. This step is clearly not necessary.

### Unequal Sample Spacing

A technique [1] for reconstructing bandlimited signals from unequally spaced samples using an iteration could be applied at the final stage to obtain greater accuracy. Due to the close spacing of the samples, this seemed unnecessary. Similarly, the FM demodulation technique based on the same iteration [2] could be applied. This also seemed unnecessary since the modulation rate is so low compared to the carrier. Under more difficult circumstances, this technique would be very useful.

### Final Results

The estimated instantaneous frequency using the reciprocal of twice each zero crossing interval is shown in Figure 4A, the noise reduction effect of using zero crossings spaced by about 0.056 s is shown in Figure 4B. The estimates at the 16 times requested by RADC are given in Table 1. These values are estimated to have an rms error of 0.7% from the error in the original estimates and the error due to the final interpolation.

825799-U

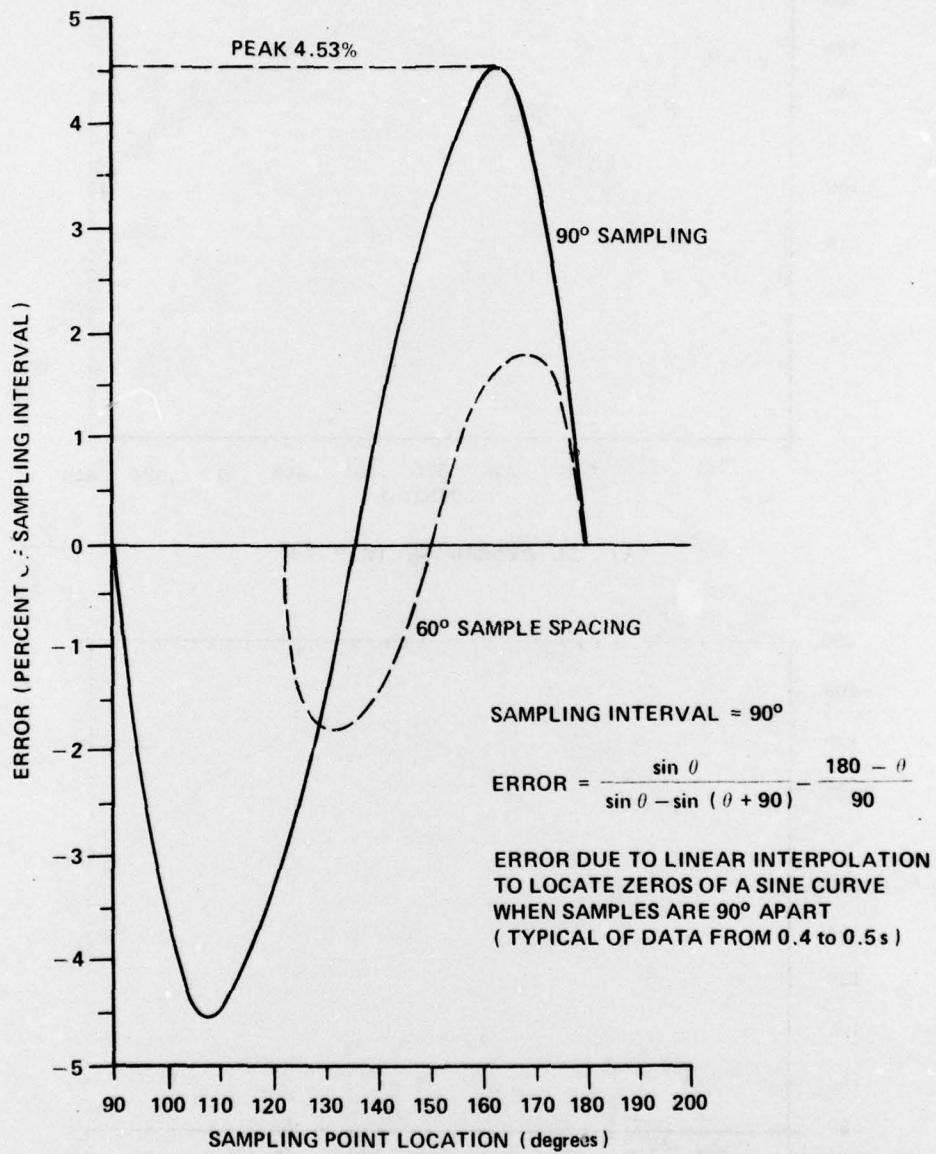
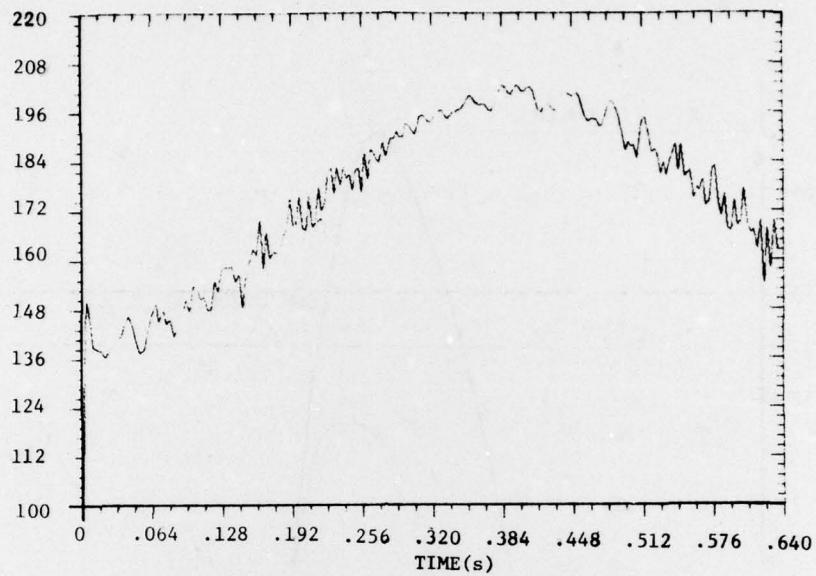
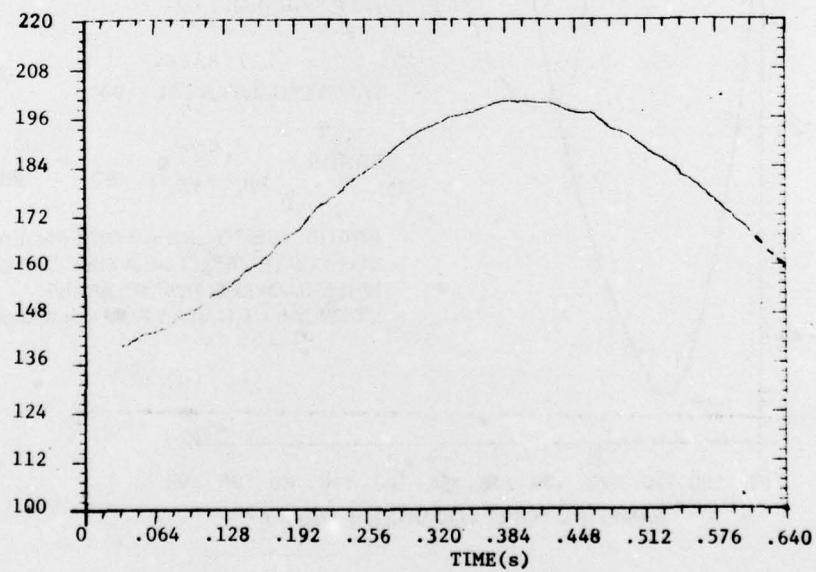


Figure 3. Error due to Linear Interpolation for Zero Location



(A) NO AVERAGING ( $K = 1$ )



(B) 0.056 s AVERAGING

Figure 4. Instantaneous Frequency Plots

Table 1. Instantaneous Frequency Estimates at 0.04 s  
Intervals

Samples	Time(s)	Instantaneous Frequency
$t_1$	0.04	140.9301
$t_2$	0.08	146.0931
$t_3$	0.12	152.9369
$t_4$	0.16	160.6346
$t_5$	0.20	169.9827
$t_6$	0.24	179.5855
$t_7$	0.28	187.5818
$t_8$	0.32	194.2402
$t_9$	0.36	197.8844
$t_{10}$	0.40	199.8585
$t_{11}$	0.44	197.8066
$t_{12}$	0.48	193.4298
$t_{13}$	0.52	186.9661
$t_{14}$	0.56	179.3949
$t_{15}$	0.60	169.5555
$t_{16}$	0.64	159.7161

## Appendix A

### Linear FM Demodulation Using Zero Crossings

#### Analysis of Zero Crossing Intervals

Assume that the FM signal is linear at least through three zero crossings, so that this segment of the FM wave is given by

$$S(t) = \sin(\omega_c t + \alpha t^2) \quad (A-1)$$

In this case,

$$\phi(t) = \omega_c t + \alpha t^2 = \text{total phase} \quad (A-2)$$

$$\theta(t) = \alpha t^2 = \text{phase modulation} \quad (A-3)$$

The instantaneous frequency is  $\frac{1}{2\pi} \dot{\phi}$ , or

$$f = \frac{1}{2\pi} \dot{\phi} = \frac{1}{2\pi} (\omega_c + 2\alpha t) \quad (A-4)$$

The slope of the modulating signal is  $2\alpha$ , and  $\omega_c$  is the carrier frequency. Clearly, the zero crossings of  $S(t)$  occur when the phase is a multiple of  $\pi$ :

$$\phi(t) = n\pi = \omega_c t + \alpha t^2 \quad (A-5)$$

where

$$n = 0, \pm 1, \pm 2, \dots \quad (A-6)$$

The values of  $t$  at which these zero crossings occur can be obtained by solving the quadratic in Equation A-5, so that

$$t_n = \frac{\omega_c}{2\alpha} \left[ \pm \sqrt{1 + \frac{4\pi n}{\omega_c^2}} - 1 \right]. \quad (A-7)$$

The zero crossing interval can be found by taking the difference between two adjacent zero crossing times:

$$\Delta t_{n+1} = t_{n+1} - t_n \quad (A-8)$$

$$\Delta t_{n+1} = \frac{\omega_c}{2\alpha} \left[ \pm \sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} - \pm \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}} \right] \quad (A-9)$$

### Computation of Instantaneous Frequency

Babbit and Leon [3] show that the average frequency over the zero cross interval is equal to the reciprocal of twice the interval, therefore:

$$f_{n+1} = \frac{1}{2\Delta t_{n+1}} \quad (A-10)$$

$$f_{n+1} = \frac{\alpha}{\omega_c} \left[ \pm \sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} \mp \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}} \right]^{-1} \quad (A-11)$$

Multiplication of Equation A-11 by

$$1 = \frac{\pm \sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} \pm \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}}}{\pm \sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} \mp \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}}} \quad (A-12)$$

simplifies the expression for frequency:

$$f_{n+1} = \frac{\omega_c}{4\pi} \left[ \sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} + \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}} \right] \quad (A-13)$$

By virtue of the mean value theorem for continuous differentiable functions, it is clear that the average value over the interval must be the actual value at some instant within the interval. Since the modulating function in question is linear, it is expected that the average given by Equation A-13 would be the instantaneous value at the midpoint of the interval. To check this, the midpoint instantaneous frequency can be computed.

$$\text{Midpoint} = \frac{t_{n+1} + t_n}{2} \quad (A-14)$$

$$\text{Midpoint} = \frac{\omega_c}{4\alpha} \left[ \sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} + \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}} - 2 \right] \quad (A-15)$$

Substituting this value of time for the midpoint given by Equation A-15 into the expression for instantaneous frequency given by Equation A-14 yields the following result:

$$f_{\text{midpoint}} = \frac{1}{2\pi} [\omega_c + \frac{\omega_c}{2} (\sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega^2}} + \sqrt{1 + \frac{4\pi\alpha(n)}{\omega_c^2} - 2})] \quad (\text{A-16})$$

$$= \frac{\omega_c}{4\pi} [\sqrt{1 + \frac{4\pi\alpha(n+1)}{\omega_c^2}} + \sqrt{1 + \frac{4\pi\alpha n}{\omega_c^2}}] \quad (\text{A-17})$$

Since Equations A-17 and A-13 are the same, the fact that the average frequency over the interval is equal to the instantaneous frequency at the midpoint of the interval for a linear FM signal has been proven.

#### References

1. R.G. Wiley, "Recovery of Bandlimited Signals from Unequally Spaced Samples" IEEE Transactions on Communications, Vol. COM-26, No. 1, January 1978, Pages 135-137.
2. R.G. Wiley, H. Schwarzlander, and D.D. Weiner, "Demodulation Procedure for Very Wideband FM", IEEE Transactions on Communications, Vol. COM-25, No. 3, March 1977, Pages 318-327.
3. Babbitt and Leon, "A Theory of FM Demodulation Based on Zero Crossings and Modified Sampling Theorems," TR-EE-70-40, December 1970, Purdue University, Lafayette, Indiana.
4. McClellan, et al., "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters," IEEE Transactions Audio, Vol. AU-21, No. 6, December 1973.
5. R.W. Schaffer and L.R. Rabiner, "A Digital Signal Processing Approach to Interpolation," Proceedings to IEEE, Vol. 61, No. 6, June 1973.